ChinaXiv合作期刊

心理科学进展 2022, Vol. 30, No. 9, 2117−2130
Advances in Psychological Science
© 2022 中国科学院心理研究所
https://doi.org/10.3724/SP.J.1042.2022.02117

chinaXiv:202303.09766v1

·研究方法(Research Method)·

结构方程模型统计检验力分析：原理与方法*

翟宏堃 李 强 魏晓薇

(南开大学社会心理学系, 天津 300350)

摘 要 结构方程模型是心理学、管理学、社会学等学科中重要的统计工具之一。然而, 大量使用结构方程模型的研究忽视了对该方法的统计检验力进行必要的分析和报告, 在一定程度上降低了这些研究的结果的证明效力。结构方程模型的统计检验力分析方法主要有 Satorra-Saris 法、MacCallum 法与 Monte Carlo 法三类。其中 Satorra-Saris 法适用于备择模型清晰、检验对象相对简单、检验方法基于 χ^2 分布的情形; MacCallum 法适用于基于 χ^2 分布的模型拟合检验且备择模型不明的情形; Monte Carlo 法适用于检验对象相对复杂、采用模拟或重抽样方法进行检验的情形。在实际应用中, 研究者应当首先判断检验的目的、方法以及是否有明确的备择模型, 并根据这些信息选择具体的分析方法。

关键词 结构方程模型, 统计检验力, 模型拟合检验, 模型参数检验
分类号 B841

1 引言

结构方程模型(Structural Equation Model, SEM)是心理学、管理学、社会学等学科中重要的统计工具之一, 可以用于验证性因子分析、共同方法偏差检验、中介/调节效应分析、交叉滞后分析等多种场景。有研究者指出, 相较于一般的回归分析, 结构方程模型能够更好地控制测量误差, 也支持构建复杂的多变量模型(王阳 等, 2020)。然而, 在使用结构方程模型进行分析时, 却较少有研究者报告统计检验力的相关内容。事实上, 统计检验力的大小在零假设显著性检验(Null Hypothesis Significance Testing, NHST)中有着重要的参考价值。

统计检验力指的是某假设检验能够正确地拒绝错误零假设的概率。对统计检验力进行分析, 可以帮助研究者确定合理样本量, 避免造成大量的人力和物力的浪费(吴艳, 温忠麟, 2011; 赵礼, 王晖, 2019)。此外, 对评估结构方程模型拟合的 χ^2 检验而言, 零假设的形式通常为模型拟合良好或模型间无差异, 而备择假设则通常为模型拟合不良或模型间存在差异。针对此类假设, Ⅱ类错误意味着研究者可能将拟合不良的模型误认为拟合良好的模型进行结果报告。王阳等人(2020)则指出, 若模型设定有误, 建立在此模型基础上的参数估计结果都是不可靠的。在这种情况下, Ⅱ类错误可能带来比Ⅰ类错误更大的危害。

近年来, 有一些研究者开始逐渐意识到使用结构方程模型作为统计工具时统计检验力分析的重要性, 并开始在研究中报告其先验或后验检验力, 或者至少在讨论部分指出该研究可能在统计检验力上存有不足(Mullen & Crowe, 2017; Hollerbach et al., 2018; Kornadt et al., 2019; Zhang & Zaman, 2020; Zhai et al., 2021)。但值得注意的是, 目前仍有许多研究者在使用结构方程模型时并未提及统计检验力的有关内容。笔者曾以全文包含“结构方程模型”或“SEM 模型”为条件, 搜索了国内心理学顶级刊物《心理学报》中近 5 年(2017~2021)内发表的论文, 共检索出 51 篇论文, 排除其中 6 篇方法类研究和 2 篇元分析研究, 余下 43 篇中提及“统计检验力”或“统计效力”的仅有 2 篇研究。本文介绍结构方程模型统计检验力分析的基本原理

收稿日期：2021-10-09
* 国家社会科学基金项目(19ASH012)资助。
通信作者：魏晓薇, E-mail: mcqueen91@163.com

ChinaXiv合作期刊

chinaXiv:202303.09766v1

和具体分析流程，并进行了实例演示，最后对相关问题进行了讨论。

2 结构方程模型中的假设检验

假设检验的技术路线直接决定了检验力分析所能选取的技术路线，结构方程模型中假设检验的技术路线大致可分为两种：基于 χ^2 分布的方法以及基于模拟(simulation)或重抽样(resampling)技术的方法。

2.1 基于 χ^2 分布的结构方程模型检验

基于 χ^2 分布的结构方程模型检验主要包括用于评价模型拟合(或嵌套模型比较)的 χ^2 检验、$\Delta\chi^2$ 检验、等效性检验、基于模型拟合指数的检验以及用于评价模型参数(或参数的函数，如中介效应)的 Wald 检验、拉格朗日乘子检验(Lagrange Multiplier test, LM 检验)和似然比检验(Satorra, 1989; Chou & Bentler, 1990; Gonzalez & Griffin, 2001; Yuan & Chan, 2016; Yuan et al., 2016; 王阳等, 2020)。这些检验涵盖了结构方程模型相关假设检验的方方面面，甚至可以说，目前与结构方程模型有关的参数检验方法要么本身就是 χ^2 检验，要么可以通过适当变形转化为 χ^2 检验。根据构建统计量时的细微差别，上述检验又可被划分为三个子类。

首先是 χ^2 检验、$\Delta\chi^2$ 检验、等效性检验和似然比检验，这四种检验都以结构方程模型的拟合函数为基础。结构方程模型的拟合函数衡量了样本协方差矩阵 $\boldsymbol{S}$ 与假设模型对应的协方差矩阵 $\boldsymbol{\Sigma}_0$ 之间的差距，亦即变量间关系在假设模型与实际数据中的差距(吴明隆，2010；王阳 等，2020)。Muthén(2004)指出，当样本来自假设模型所对应的正态总体时，拟合函数值的 $N-1$ 倍(N 为样本量)随着 N 的增大渐近服从中心化 χ^2 分布。另有学者则指出，当假定样本来自与假设模型差距小于给定值(记作 F_0)的正态总体时，此时拟合函数值的 $N-1$ 倍渐进服从非中心参数为$(N-1)F_0$ 的非中心 χ^2 分布(Satorra & Saris, 1985; MacCallum et al., 1996; Yuan & Chan, 2016)。据此，可以构建出零假设成立时 χ^2 检验和等效性检验所对应的统计量分布。而 $\Delta\chi^2$ 检验和似然比检验则是在前者的基础上，增加了对假设模型的约束条件，通过比较添加约束条件前后实际数据与假设模型的差距是否显著增大来判定添加的约束条件是否“合理”。

第二类是基于模型拟合指数的检验。从原理上讲，这类检验与第一类检验差别不大，其本质是通过拟合指数的定义式将拟合指数换算为非中心参数，因此该方法仅能用于根据非中心参数定义的拟合指数(如 RMSEA、CFI、MFI、GFI[1] 等)。此类检验有两种相似但略有不同的技术路线，一种将基于模型拟合指数的零假设换算为基于非中心参数的零假设，选择非中心参数的临界值作为构建 χ^2 统计量分布的依据, MacCallum 等人(1996)和 MacCallum 等人(2006)就采用了这种做法。另一种则根据 Venables (1975)的研究，以样本计算出的非中心参数估计值为依据构建总体非中心参数的 $1-\alpha$ 置信区间，再根据定义式将其换算回拟合指数，从而得到拟合指数的 $1-\alpha$ 置信区间，进而完成统计推断。基于 CFI 与 RMSEA 的等效性检验则采用了此类做法(Yuan et al., 2016; Marcoulides & Yuan, 2017; 王阳 等, 2020)。

第三类是 Wald 检验和 LM 检验。与前两类方法不同的是，此类方法不再关注样本的拟合函数值本身(非中心参数也是关于拟合函数的函数)，而是通过对关键参数及其标准误的估计，构建 z 统计量进行检验。其中 Wald 检验以备择假设成立时的模型为基础完成对目标参数的估计和检验。当处理参数函数形式的假设时，Wald 检验对参数函数在其真值邻域内进行一阶 Taylor 展开求得其线性近似值，进而完成对待检验指标的标准误估计(Chou & Bentler, 1990; Gonzalez & Griffin, 2001)。此时 Wald 检验亦称为 Delta 法，叶宝娟和温忠麟(2012)曾介绍过如何利用该方法对多维测验合成信度的置信区间进行估计。与 Wald 检验相对，LM 检验以零假设成立时的模型为基础，将假设条件重写为约束函数的形式，并构建拉格朗日函数将原假设检验问题变为检验拉格朗日乘数是否为 0 的问题(Chou & Bentler, 1990)。若拉格朗日乘数为 0，则意味着新假设条件的加入没有改变原假设模型。

2.2 基于模拟或重抽样技术的结构方程模型检验

近年来，随着计算机技术的发展，模拟技术和重抽样技术越来越多地出现在统计学领域，这

[1] 本文中出现的“GFI”和“AGFI”指的分别是“Unbiased goodness of fit index”和“Unbiased adjusted goodness of fit index”，而非“Goodness of fit index”和“Adjusted goodness of fit index”，详见网络版补充材料。

ChinaXiv合作期刊

chinaXiv:202303.09766v1

些技术被广泛地运用在包括参数估计与假设检验在内的多种场合。不同于基于理论分布的经典方法，此类方法通常以模拟或重抽样的方式得到目标统计量的经验分布，并以经验分布代替理论分布进行统计推断，因此往往用于目标统计量的理论分布不明确或难于计算的情形。此类方法中最为重要的两种分别是基于模拟技术的 Monte Carlo (MC)方法[2]与基于重抽样技术的 Bootstrap 方法。

MC 方法最早由 Stanisław Ulam 等人于 20 世纪 40 年代中期提出，其核心思想为采用模拟的方式“重现”某一概率模型所描绘的过程，以模拟得到的结果作为目标问题的近似解。MacKinnon 等人(2004)在评估中介效应检验的 M 法和经验 M 法时同时引入了 MC 法。Preacher 和 Selig (2012)进一步指出 MC 法不仅可以用于线性回归和路径分析模型，也可直接扩展至潜变量模型。该方法假定模型参数服从联合正态分布，其均值向量与协方差矩阵通常由对原始模型的极大似然估计确定。MC 法将模型参数视为从确定的联合正态分布中抽样的过程，以此为基础进行随机抽样，最终基于抽样结果构建中介效应的经验分布并进行统计推断(Preacher & Selig, 2012)。此外，在结构方程建模中还有一类 Monte Carlo 方法——马尔可夫链蒙特卡洛(Markov chain Monte Carlo, MCMC)方法，此类方法主要用于贝叶斯结构方程建模中。由于对“参数”这一重要概念的基础假设不同，贝叶斯统计中假设检验和统计检验力分析的过程均与基于频率学派的传统统计学有着较大的区别，本文的内容仍然在频率学派的框架内，故对该方法不做过多介绍。

注意到，基于模拟技术的结构方程模型检验仍然对分布有着较强的假设，MC 法假定路径系数服从联合正态分布，基于贝叶斯统计的 MCMC 法需要给出模型参数的先验分布与给定参数时样本的条件分布。相比之下，基于重抽样技术的检验对样本分布的假设更弱，Efron (1979)提出的 Bootstrap 法[3]仅要求样本是从某一确定总体中随机抽样得到的，对其分布形态并无更多要求。不仅如此，Bootstrap 法同样基于“利用经验分布代替理论分布进行统计推断”的思路，在面对一些理论分布不明或难以计算的场合表现良好。由于 Bootstrap 法具有前述技术优势，近年来，该方法已逐渐成为结构方程模型假设检验问题中的显学，被广泛应用于多种不同场合。Zhang 和 Savalei (2016)指出可以用 Bootstrap 法估计结构方程模型拟合指数的置信区间。Rosseel 在其编写的 R 程序包 lavaan 中给出了求结构方程模型路径系数的 Bootstrap 标准误的程序(Rosseel, 2021)。更有研究者利用 Bootstrap 法估计一些根据结构方程模型路径系数计算得到的更复杂的指标。屠金路等人(2005)利用 Bootstrap 法对问卷合成信度的置信区间进行了估计；张涵和康飞(2016)介绍了基于 Bootstrap 的多重中介效应分析。方杰等人(2011)的研究及温忠麟和叶宝娟(2014a, 2014b)的研究则分别介绍了 Bootstrap 法在中介效应与有调节的中介效应检验中的应用。

表 1 中按照不同类别总结了结构方程模型涉及到的常见检验的应用场景、零假设、统计量构建方式、零假设成立时统计量对应的分布以及使用中的优缺点。

3 结构方程模型假设检验统计检验力的分析方法

统计检验力的分析方法与分析对象所采用的技术路径息息相关，结构方程模型假设检验统计检验力的分析方法也可大致分为两条不同的技术路线：基于 χ^2 分布的方法和基于模拟的方法。其中，基于 χ^2 分布的方法可以用于解决使用经典方法进行结构方程模型假设检验时的统计检验力分析问题，而基于模拟的方法可以用于解决一般的结构方程模型假设检验的统计检验力分析问题。

3.1 基于 χ^2 分布的统计检验力分析

顾名思义，基于 χ^2 分布的方法以 χ^2 分布的相关理论为基础。此类方法可以追溯到 Satorra 和

[2] Monte Carlo 方法实际上代表的是以计算机模拟为核心的一组方法，为了和后文统计检验力分析的 Monte Carlo 方法加以区分，假设检验的 Monte Carlo 方法本文统一简称为 MC 方法/MC 法，特此说明。

[3] 事实上，Bootstrap 法还分为参数 Bootstrap 法和非参数 Bootstrap 法，参数 Bootstrap 法由于需要了解随机变量的分布型态，在实际中使用不多。本文中若不加说明，Bootstrap 法特指非参数 Bootstrap 法。

表 1　结构方程模型中的假设检验

检验方法所处的类别			具体的检验方法	应用场景	H_0	统计量构成	零假设成立时统计量对应的分布	优点	缺点
基于 χ^2 分布的方法(经典方法)	χ^2 检验	χ^2 检验	模型拟合的 χ^2 检验	检验模型拟合	$F_{ML0}=0$ $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$	$(n-1)\cdot F_{ML}$	中心化 χ^2 分布	计算简便；工具成熟	对样本的分布有较强的假设；对统计量的构成有一定限制；面对较为复杂的参数函数(如中介效应)表现不佳
			基于 χ^2 统计量的等效性检验	检验模型拟合	$F_{ML0}>\varepsilon$ $\boldsymbol{\Sigma}$ 与 $\boldsymbol{\Sigma}_0$ 间的差距大于 ε	$(n-1)\cdot F_{ML}$	非中心 χ^2 分布		
		$\Delta\chi^2$ 检验	嵌套模型比较的 $\Delta\chi^2$ 检验	比较模型优劣	$F_{ML0B}-F_{ML0A}=0$ $\boldsymbol{\Sigma}_{0A}=\boldsymbol{\Sigma}_{0B}$	$(n-1)\cdot(F_{MLB}-F_{MLA})$	中心化 χ^2 分布		
			模型参数的似然比检验	检验特定参数	$g(\theta)=0$ $\boldsymbol{\Sigma}_{0\,约束}=\boldsymbol{\Sigma}_{0\,无约束}$	$(n-1)\cdot(F_{ML\,约束}-F_{ML\,无约束})$	中心化 χ^2 分布		
	基于 χ^2 分布理论的检验	Z 检验	Wald 检验	检验特定参数	$g(\theta)=0$ $\boldsymbol{\Sigma}_{0\,约束}=\boldsymbol{\Sigma}_{0\,无约束}$	$\frac{g(\theta)}{\mathrm{SE}_{g(\theta)}}$	标准正态分布		
			LM 检验	检验特定参数	$\lambda=0$ $\boldsymbol{\Sigma}_{0\,约束}=\boldsymbol{\Sigma}_{0\,无约束}$	$\frac{\lambda}{\mathrm{SE}_{\lambda}}$	标准正态分布		
		χ^2 分布非中心参数检验	拟合指数的检验	检验模型拟合	$\delta=0$ $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$	给定的拟合指数	不明，但临界值与中心化 χ^2 分布有关		
			基于拟合指数的等效性检验	检验模型拟合	$\delta>\delta_k$ $\boldsymbol{\Sigma}$ 与 $\boldsymbol{\Sigma}_0$ 间的差距大于 $f(\delta_k)$	给定的拟合指数	不明，但临界值与非中心 χ^2 分布有关		
基于模拟或重抽样技术的方法	基于模拟技术的方法	Monte Carlo 法	MC 法	检验模型拟合	$F_{ML0}=0$ $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$	给定的拟合指数(包括 χ^2 统计量)	经验分布	放宽了对样本的假设；不需讨论目标统计量的理论分布	计算较为困难、计算耗时较长、部分方法要求使用者有编程技术
				检验特定参数	$g(\theta)=0$ $\boldsymbol{\Sigma}_{0\,约束}=\boldsymbol{\Sigma}_{0\,无约束}$	$g(\theta)$	经验分布		
	基于重抽样技术的方法	Bootstrap 法	Bootstrap 法	检验模型拟合	$F_{ML0}=0$ $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$	给定的拟合指数(包括 χ^2 统计量)	经验分布	灵活度高，对统计量的构造几乎没有限制；对样本分布要求宽松	
				检验特定参数	$g(\theta)=0$ $\boldsymbol{\Sigma}_{0\,约束}=\boldsymbol{\Sigma}_{0\,无约束}$	$g(\theta)$	经验分布		

Saris 早年的一系列研究(Satorra & Saris, 1985; Satorra, 1989; Satorra et al., 1991)。Satorra (1989)指出，当前述假设检验的 H_0 不成立时，其对应统计量(对 Wald 检验和 LM 检验而言是统计量的平方)服从非中心 χ^2 分布，且其相应的非中心参数可以通过总体协方差矩阵真值拟合零假设模型计算出的统计量得到。进而可以通过计算备择假设成立条件下目标统计量落入拒绝域的概率求得给定检验的统计检验力。设零假设和备择假设成立时对应的非中心参数分别为 δ_0 和 δ_1，则统计检验力 $1-\beta$ 由下式给出：

$$1-\beta=P\left\{\chi^2_{df,\delta_1}<T_{MLC}\right\}$$

(拒绝域在左侧, 对应等效性检验)　(1a)

$$\text{或}1-\beta=P\left\{\chi^2_{df,\delta_1}>T_{MLC}\right\}\text{拒绝域在右侧,}$$

(对应除等效性检验外的其他检验)　(1b)

其中 $\chi^2_{df,\delta}$ 表示自由度为 df，非中心参数为 δ 的非中心卡方分布，T_{MLC} 由下式确定：

$$P\left\{\chi^2_{df,\delta_0}\geqslant T_{MLC}\right\}=\alpha$$

(拒绝域在左侧, 对应等效性检验)　(2a)

$$\text{或}P\left\{\chi^2_{df,\delta_0}\leqslant T_{MLC}\right\}=\alpha\text{(拒绝域在右侧,}$$

对应除等效性检验外的其他检验)　(2b)

ChinaXiv合作期刊

chinaXiv:202303.09766v1

由于 δ_0 在 H_0 确定时已经可以算出, 因此, 基于 χ^2 分布的统计检验力分析的关键在于估计 δ_1。根据估计非中心参数的方法不同, 此类方法又可分为利用总体协方差矩阵 **Σ** 进行估计的 Satorra-Saris 法和利用拟合指数(如 RMSEA、CFI、MFI 等)进行估计的 MacCallum 法。

3.1.1　Satorra-Saris 法

Satorra-Saris 法主要基于 Satorra 和 Saris 先前的研究, 其具体步骤如下(Satorra & Saris, 1985; Satorra et al., 1991):

第一步: 定义备择假设成立情况下的模型, 指定备择假设成立情况下的理论总体协方差矩阵 **Σ**。

第二步: 设定样本容量为真实样本容量 *N*, 利用 **Σ** 拟合零假设模型。

第三步: 将第二步中计算得到的统计量(针对 χ^2 分布的统计量为统计量本身, 针对标准正态分布的统计量为统计量的平方)作为 δ_1 的估计, 代入(1a)或(1b)完成统计检验力计算。

使用 Satorra-Saris 法估计达到某统计检验力所需样本量的过程与估计统计检验力类似, 只是第二步中需要反复设定样本容量 *N* 进行多次估计, 并计算不同样本容量对应的统计检验力, 最终从中选取满足目标统计检验力所需的最小样本容量。

Satorra-Saris 法通过定义不同于零假设模型的备择模型重构总体协方差矩阵, 并据此计算统计检验力。其中, 对备择模型进行定义的过程不仅需要研究者给出明确的模型结构, 还需要研究者指定大量的参数(如路径系数、测量残差等)。因此, 从应用的角度考虑, Satorra-Saris 法更适合在研究者对备择模型的了解程度较深的情况下使用(如针对模型参数的检验)。

3.1.2　MacCallum 法

Satorra-Saris 法对备择模型过强的假设从客观上提高了该方法的使用难度, 也限制了其适用范围。MacCallum 等人(1996)给出了另外一种估计非中心参数的方法, 该方法同样基于 Satorra 和 Saris (1985)给出的公式, 但在估计 F_{ML0} 时, MacCallum 等人给出了一个基于 RMSEA 的估计方法。他们指出, 使用样本拟合函数估计值 F_{ML} 作为对总体参数 F_{ML0} 的估计是有偏的, 且 F_{ML0} 与 F_{ML} 之间有如下关系(MacCallum et al., 1996):

$$F_{\mathrm{ML0}} = F_{\mathrm{ML}} - \frac{df}{N-1} = df \cdot \mathrm{RMSEA}^2 \qquad (3)$$

据此, MacCallum 等人提出基于 RMSEA 指标的统计检验力分析步骤(MacCallum et al., 1996):

第一步: 根据零假设和备择假设成立时对应的 RMSEA 值、模型自由度以及样本量分别计算出零假设和备择假设成立时的非中心参数 δ_0 和 δ_1。

第二步: 代入(1a)或(1b)完成统计检验力计算。

与 Satorra-Saris 法类似, 当使用 MacCallum 法估计达到某统计检验力所需样本量时, 也只需要在第一步中反复设定不同的 *N* 对统计检验力进行多次估计, 最终从中选取满足目标统计检验力所需的最小样本容量。

Kim (2005)对 MacCallum 法进行了进一步简化和扩展, 给出了一个快速求取给定统计检验力情况下所需样本量的公式:

$$N_{RMSEA} = \frac{\delta}{df \cdot \mathrm{RMSEA}^2} + 1 \qquad (4)$$

其中 *df* 为模型自由度, RMSEA 为备择假设成立时对应的 RMSEA 值, δ 为显著性水平为 α、统计检验力为 1−β 情况下对应的非中心参数。非中心参数可以通过查表或者使用统计软件计算得到, Kim (2005)在他们的文章最后同时给出了 H_0: $F_{ML0} = 0$ 时利用计算机快速求取 δ 的 SPSS 与 SAS 脚本。此外, Kim (2005)在他的文章中还提出了基于 CFI、MFI 与 GFI 的计算公式。

相比于 Satorra-Saris 法, MacCallum 法采用拟合指数对模型的拟合情况进行了某种意义上的"打包", 即利用拟合指数对假设模型与实际数据的差距进行了概括, 使研究者免于设定复杂的备择模型, 在分析模型拟合检验的统计检验力时, MacCallum 法仅需要模型自由度、零假设和备择假设成立时对应的拟合指数值[若对关于 CFI 的检验进行检验力分析时还需要提供基础模型(base model)的拟合函数值 F_{MLB} 与自由度 df_B, 关于 GFI 的检验还需要提供协方差矩阵的秩 *p*], 大大简化了统计检验力的分析难度。但由于模型参数与拟合指数之间存在复杂的非线性关系, MacCallum 法无法分析模型参数检验的统计检验力。

3.2　基于模拟的统计检验力分析

采用模拟方法对结构方程模型假设检验进行统计检验力分析早年可追溯到 Muthén 与 Muthén (2002)的研究。采用模拟方法进行统计检验力分

ChinaXiv合作期刊

析的思路与采用模拟方法进行结构方程模型假设检验的思路相似，即利用模拟的方式确定$P\left(\text{拒绝}H_0 \mid \theta=\theta^*\right)$的经验值。具体而言，此类方法通常采用模拟的方式从参数为 θ^*的特定分布中生成多组随机样本，而后针对每组随机样本执行同样的假设检验程序，最后统计全部已进行的假设检验中拒绝零假设的频率，以此作为对该检验统计检验力的分析结果。Muthén 与 Muthén (2002)的研究给出了利用 Monte Carlo 模拟计算基于经典方法的结构方程模型假设检验的统计检验力的分析过程与 Mplus 程序。Zhang (2014)的研究指出 Monte Carlo 模拟法同样可以估计基于 Bootstrap 方法的中介效应检验的统计检验力，并给出了一个基于 R 语言 bmem 程序包的示例程序。Schoemann 等人(2017)的研究在实现对中介模型的统计检验力分析时使用了 Monte Carlo 模拟法估计基于 MC 法的中介效应检验的统计检验力。总的来说，使用 Monte Carlo 模拟法进行统计检验力分析的步骤大体如下(Muthén & Muthén, 2002; Zhang, 2014; Schoemann et al., 2017):

第一步：根据备择假设成立时的情况设定用于生成模拟数据的均值向量(模型不考虑均值结构时为零向量)与协方差矩阵。此步骤中既可以如 Schoemann 等人(2017)直接指定协方差矩阵，也可如 Zhang (2014)通过定义模型来间接指定协方差矩阵。此外，如果研究者希望模拟非正态分布的数据，通常还需要给出目标分布的峰度和偏度。

第二步：确定需要进行 Monte Carlo 模拟的重复轮数 R 与样本容量 N。Zhang (2014)建议重复轮数 R 通常不应少于 1000。

第三步：在每一轮 Monte Carlo 模拟中，重复执行下述步骤：(1)根据给定的分布参数与样本量生成服从特定分布的随机数作为模拟样本；(2)对该模拟样本进行结构方程建模并进行检验，具体的检验方法可以是表 1 中的任意一种。

第四步：统计出每一轮统计检验的结果，设在 R 轮检验中有 r 轮结果拒绝了原假设，则最终的统计检验力为 $\frac{r}{R}$。

此外，为了说明 Monte Carlo 模拟结果的准确性，Muthén 和 Muthén (2002)建议研究者在报告结果的同时报告结构方程模型中各参数的理论真值、参数估计结果的平均值和标准差、参数估计标准误的平均值以及参数 95%置信区间对理论真值的覆盖率。Muthén 和 Muthén (2002)进一步提出衡量 Monte Carlo 模拟结果是否可靠的标准：(1)模型中任何参数及其标准误真值与估计值之间的相对误差不超过 10% (参数估计结果的标准差视为标准误的真值)；(2)模型中关键参数及其标准误真值与估计值之间的相对误差不超过 5%；(3)参数的 95%置信区间对理论真值的覆盖率应当在 0.91 至 0.98 之间。

当使用 Monte Carlo 法估计达到某统计检验力所需样本量时，也只需要在第一步中反复设定不同的 N 对统计检验力进行多次估计，而后通过线性插值法对所需样本量进行估计。与基于 χ^2 分布的统计检验力分析方法相比，结构方程模型统计检验力的 Monte Carlo 分析由于不需要计算目标统计量的理论分布，因此可以广泛适用于对各种检验进行统计检验力分析的情形，在给出峰度和偏度时，该方法还可以在一定程度上分析总体分布为非正态分布的情形。但结构方程模型统计检验力分析的 Monte Carlo 法需要研究者至少给出用以生成模拟数据的协方差矩阵，这意味着研究者需要对备择假设成立下数据的分布情况有足够的认识。另外，该方法本身对研究者的编程水平要求较高，且计算量较大，计算起来较为耗时，在一定程度上影响了该方法的使用和推广。

3.3 结构方程模型统计检验力分析的程序资源

为了方便研究者对统计检验力进行分析，多名研究者提供了在线或离线的相关程序。基于 χ^2 分布的统计检验力分析方法中，Preacher 和 Coffman (2006)提供了一个基于 MacCallum 法的统计检验力在线计算程序，可以计算给定样本量情况下，单个模型拟合或嵌套模型拟合差异的统计效力或达到某统计效力所需的样本量。Moshagen 提供了一个 R 程序包 semPower，该程序包基于 Moshagen 和 Erdfelder (2016)的研究，可以计算基于 F_{ML0}、RMSEA、McDonald 拟合指数、GFI 与 AGFI 的统计检验力，当指定零假设和备择假设成立情况下对应的协方差矩阵时，该程序包也可使用 Satorra-Saris 法分析统计检验力。Jak 等人(2021)给出了一个基于 Shiny 开发的在线程序 power4SEM (网址为 https://sjak.shinyapps.io/power4SEM/)，该程序同时支持 Satorra-Saris 法和 MacCallum 法，

chinaXiv:202303.09766v1

ChinaXiv合作期刊

Jak 等人在他们的论文中详细介绍了该程序的使用方式。

基于模拟的统计检验力分析方法中，Muthén 和 Muthén (2002)提供了使用 Monte Carlo 法分析结构方程模型统计检验力的 Mplus 程序框架。Zhang 和 Wang (2020)提供了可以用来分析中介效应统计检验力的 R 语言程序包 bmem。该程序包使用 Monte Carlo 法进行统计检验力分析，可以对中介效应的 Sobel 检验与 Bootstrap 检验的统计检验力进行分析。特别的是，该程序包还提供给研究者自主设定分布的峰度与偏度的功能以模拟总体为非正态分布时的情形。Schoemann 等人(2017)给出了一个使用 Monte Carlo 法分析中介效应统计检验力的在线程序(网址为 https://schoemanna.shinyapps.io/mc_power_med/)，该程序基于 Shiny 开发，可用来处理简单中介模型、两个中介变量的平行中介模型和两个中介变量的链式中介模型三种情况下的中介效应统计检验力分析。Wang 和 Rhemtulla (2021)则给出了一个分析结构方程模型参数检验统计检验力的在线程序(网址为: https://yilinandrewang.shinyapps.io/pwrSEM/)。

表 2 中总结了结构方程模型统计检验力分析方法的适用场合、需要的计算量以及相应的程序资源。

4　应用实例

下面用一个例子演示在实际研究中如何灵活使用各种方法对结构方程模型的统计检验力进行分析。注意到，统计检验力的分析包括先验分析和后验分析，二者对应研究过程中的不同时期，且不能相互替代(吴艳，温忠麟，2011；赵礼，王晖，2019)。但先验分析与后验分析在分析步骤上差异不大，且亦有研究者指出，在 NHST 中，统计检验力是 p 值的函数，在已知 p 值的情况下统计检验力没有给出额外信息，因此建议无需进行后验分析(Hoenig & Heisey, 2001)。考虑到篇幅所限，本文仅展示先验分析的过程，后验分析的部分参见网络版补充材料。演示的样例来自 Zhai 等人(2021)的研究，该研究调查了新冠疫情期间个体的心理健康与创伤后成长的基本情况，并讨论了在控制疫情相关压力和年龄的情况下，情绪创造力对个体心理健康的保护作用以及对个体创伤后成长的促进作用和可能的作用机制。具体的模型设置参见图 1，统计软件采用 R 3.6.3 的 lavaan 程序包和 semPower 程序包，本部分全部 R 代码参见网络版补充材料部分。

从检验目的的角度出发，使用结构方程模型进行验证性因子分析的主要目的是通过对模型拟合的评价考察样本数据是否符合量表的理论结构，即考察量表题项与测量因子间的关系是否同研究者的假设所一致；而使用结构方程模型进行中介效应分析，特别是对显变量链式中介模型进行中介效应分析时，研究者的目的则更多指向对路径系数函数(即中介效应)的分析与检验。此外，注意到图 1 中的模型为饱和模型，有研究者指出，此时模型为完美拟合，可不关注模型拟合指数，仅讨论路径系数即可(张莉 等，2019)。综合考虑检验目的与实际情况，对该研究的统计检验力进行分析时，需要对验证性因子分析部分模型拟合检验的统计检验力与链式中介模型部分模型参数检验的统计检验力进行分析，链式中介效应部分选择“情绪创造力→领悟社会支持→情绪调节自我效能感→压力后成长”路径的中介效应值为分析对象。

表 2　结构方程模型统计检验力分析方法

分析方法	是否需要给出协方差矩阵	对应假设检验的技术路线	假设检验的目的	计算量	程序资源
Satorra-Saris 法	是	仅经典方法	均可，但更适合模型参数检验或嵌套模型比较	较小	power4SEM semPower
MacCallum 法	否	仅经典方法	模型拟合检验	最小	power4SEM semPower Preacher 在线程序
Monte Carlo 法	是	均可	均可	大	Bmem mc_power_med pwrSEM

chinaXiv:202303.09766v1

ChinaXiv合作期刊

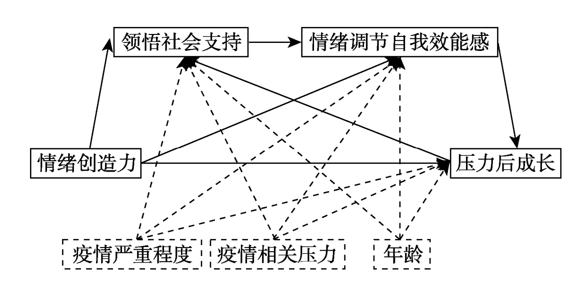


图 1 样例的结构方程模型示意图
(虚线表示该变量为控制变量)

4.1 模型拟合检验的先验检验力分析

4.1.1 使用 Satorra-Saris 法分析模型拟合检验的先验检验力

在进行先验检验力分析时，首先要设定研究对象理论上的总体协方差矩阵(下简称理论协方差矩阵)，即备择假设成立时对应的协方差矩阵。与 *t* 检验、方差分析的先验检验力分析类似，在结构方程模型相关检验的先验检验力分析中，理论协方差矩阵也可以通过参考前期理论、已发表的类似成果或研究者提出的标准所确定(吴艳，温忠麟, 2011)。通常情况下，建议研究者优先参考已有类似研究中报告的结果设定理论协方差矩阵，当缺乏既往研究或既往研究信息不充足时，研究者可以参考前期理论或其他研究者提出的标准所确定。此外，Jak 等人(2021)建议研究者在设定理论协方差矩阵时，可以以标准化的形式指定参数值，并且参考相关研究领域中“小、中、大”所对应的经验值代入待定参数。同时，由于通常情况下确定协方差矩阵所需要的参数数量较为庞大，Jak 等人(2021)还建议研究者在设定理论协方差矩阵时需要考虑那些在理论上能够影响拒绝或接受 H_0 的重要参数，例如路径系数、误差相关、潜变量相关等。

通过查询资料发现：既往研究中，文书锋等人(2009)的研究较为全面地报告了可以用于生成理论协方差矩阵的全部参数，且该研究样本量较大(验证性因子分析部分 N = 818)。因此，在之后的分析中，以该模型为基准生成理论协方差矩阵。本例中控制 I 类错误水平 α 为 0.05，求得统计检验力为 0.95 时对应的样本量约为 134 人。此外，对不熟悉 R 的研究者，也可以使用常用的结构方程建模软件(如 LISREL、EQS、AMOS、Mplus 等)结合 G*Power 进行分析，具体的步骤参见网络版补充材料部分。

4.1.2 使用 MacCallum 法分析模型拟合检验的先验检验力

相比 Satorra-Saris 法，使用 MacCallum 法分析模型拟合检验先验检验力的过程要简单许多，只需要确定零假设和备择假设成立情况下对应的拟合指数、模型自由度 *df*、显著性水平 α 和统计检验力 1 − β 五个指标即可完成分析。在设定备择假设对应的拟合指数时，研究者可参考既往研究中提出的关于该指数“大小”的标准。例如对 RMSEA 而言，有研究者指出，RMSEA 小于 0.05 为良好拟合，0.05 至 0.08 为合理拟合，0.08 至 0.1 为可接受的拟合，大于 0.1 为拟合较差(MacCallum et al., 1996; Yuan & Chan, 2016; Jak et al., 2021)。对于本文中给出的例子，控制 I 类错误水平 α 为 0.05，若研究者若希望在 RMSEA 的真值为 0.08 时，正确拒绝该模型的几率不小于 95%，此时使用 semPower 包提供的 semPower.aPriori 函数可以计算得出需要的样本量为 134 人。

4.1.3 使用 Monte Carlo 法分析模型拟合检验的先验检验力

使用 Monte Carlo 法分析模型拟合检验先验检验力时确定理论协方差矩阵的步骤和原则同 Satorra-Saris 法一致，这里不再赘述。本例中，以等效性检验为分析对象，控制 I 类错误水平 α 为 0.05，重复次数 R 为 1000。分别取 N 为 101, 201, 301 的情形，计算在给定 N 的情况下 1000 次重复试验中模型拟合指数落在以 $RMSEA_0 = 0.08$ 为经验临界值的等效性检验拒绝域的比率。需要注意的是，在 Monte Carlo 模拟的过程中，有些样本会导致结构方程模型在计算过程中出现迭代不收敛、残差方差为负数、潜变量协方差矩阵非正定等“不合理”的情况，即出现了不恰当解，这些解应在在后续的计算过程中予以排除。本例中，N 分别为 101 和 201 时对应的恰当解比例分别为 99.6%和 100%，对应的统计检验力分别为 0.724 和 0.996；对应的各参数估计情况、相对误差以及 95%置信区间覆盖率参见网络版补充材料。本例中全部参数均值和标准误的估计误差均不超过 10%, 95%置信区间覆盖率均处于 0.91 至 0.98 之间，提示 Monte Carlo 模拟结果的可信度较高。进一步根据线性插值法可求得统计检验力为 0.95 时对应的样本量约为 185 人。针对同一个假设检验进行统计检验力分析时，Monte Carlo 方法得到的结果

chinaXiv:202303.09766v1

ChinaXiv合作期刊

看似与 Satorra-Saris 法有较大差异，但这主要是由于线性插值带来的误差。事实上，若通过缩减采样点间距的方式提高线性插值的精度，可重新求得待估样本量约为 151 人。

4.2　模型参数检验的先验检验力分析

4.2.1　使用 Satorra-Saris 法分析模型参数检验的先验检验力

使用 Satorra-Saris 法分析模型参数检验先验检验力的第一步仍然是设定理论协方差矩阵。诚然经典的 Satorra-Saris 法强调通过备择模型生成理论协方差矩阵，但 Satorra 和 Saris (1985)给出的关于该方法原理的讨论中并没有限制理论协方差矩阵的生成模式。因此，参照 Schoemann 等人(2017)的方式直接设定理论协方差矩阵也并无不可。

由于既往研究中缺乏可以用来直接重构理论协方差矩阵的数据，而通过元分析技术虽然能够在一定程度上基于现有研究构建出理论协方差矩阵，但这种方法需要耗费较多的人力物力，单纯用来进行统计检验力分析未免有些浪费。因此，本文在进行理论协方差矩阵生成时，选择根据相关研究领域中“小、中、大”所对应的经验值进行设定。此外，由于尺度变换不影响对本例中所关心的系数(显变量间的路径系数)显著性的判定。因此，在指定理论协方差矩阵时可进一步将其简化为指定相关矩阵。参考 Cohen (1988)的观点，相关系数在 0.1、0.3、0.5 时分别对应 “小”、“中”、“大”的效应量。在本例中，假定与控制变量有关的相关系数有“小”的效应量，关注的变量的相关系数有“中”的效应量，以此设定理论协方差矩阵。以对“情绪创造力→领悟社会支持→情绪调节自我效能感→压力后成长”这一路径中介效应的 Wald 检验为分析对象，控制 I 类错误水平 α 为 0.05，求得统计检验力为 0.95 时对应的样本量约为 795 人。

4.2.2　使用 Monte Carlo 法分析模型参数检验的先验检验力

本例中，以对“情绪创造力→领悟社会支持→情绪调节自我效能感→压力后成长”这一路径中介效应的 Wald 检验为分析对象，控制 I 类错误水平 α 为 0.05，分别取 N 为 301, 401, 501, 601 的情形进行模拟，得到的结果全部为恰当解，对应的统计检验力为 0.613, 0.884, 0.960, 0.985，对应的各参数估计情况、相对误差以及 95%置信区间覆盖率参见网络版补充材料。本例中全部参数的估计误差均不超过 10%，关键参数的估计误差均不超过 5%，95%置信区间覆盖率均处于 0.91 至 0.98 之间，提示 Monte Carlo 模拟结果的可信度较高。进一步根据线性插值法求得统计检验力为 0.95 时对应的样本量约为 488 人。

5　讨论

5.1　先验检验力分析时的参数设定问题

注意到，研究者在使用 Monte Carlo 法或 Satorra-Saris 法进行先验检验力分析时，均需要指定理论协方差矩阵。该步骤中，不论研究者是采用直接设定协方差矩阵的方式还是通过设定理论模型进而生成协方差矩阵的方式，都需要设定大量参数，这些参数的设定会对结果产生一定影响，且影响规律较为复杂。例如，有研究要对一个两维度，每个维度包含 4 个题目的量表进行验证性因子分析。假设 H_0 为两维度间不存在相关，H_1 成立的情况下两维度之间存在相关系数为 0.3 的正相关，进行先验检验力分析时，当研究者设定载荷为 0.7、0.75、0.8 时，设定显著性水平为 0.05，统计检验力为 0.95，使用 Satorra-Saris 法计算得到的计划样本量分别为 527 人、472 人和 429 人，使用 Monte Carlo 法计算得到的计划样本量分别为 521 人、464 人和 417 人。又如，本文 4.2 部分给出的例子中，若将所有控制变量与其他变量的相关系数从 0.2 变为 0.1，在其他条件不变的情况下，使用 Satorra-Saris 法计算得到的计划样本量为 352 人，使用 Monte Carlo 法计算得到的计划样本量为 233 人。而当分析对象是中介效应或有调节的中介效应这类由参数函数(通常是乘积)构建的指标时，若研究者使用理论模型生成协方差矩阵，即使是同样大小的指标值，不同的参数也会导致分析结果的不同。例如，对简单中介效应模型而言，设定自变量到因变量的路径系数为 0.2，中介效应值为 0.16，显著性水平为 0.05，统计检验力为 0.90，中介前半路径系数(自变量到中介变量的路径系数)和中介后半路径系数(中介变量到因变量的路径系数)分别取 0.2 和 0.8，0.4 和 0.4，0.8 和 0.2(所有系数均为标准化系数)的情况下，使用 Satorra-Saris 法计算得到的计划样本量分别为 258 人、114 人和 632 人，使用 Monte Carlo 法计算得到的计划样本量分别为 255 人、88 人和 620 人。

chinaXiv:202303.09766v1

ChinaXiv合作期刊

chinaXiv:202303.09766v1

从上述例子中不难看出先验检验力分析中参数设定对结果的影响。这意味着，研究者在使用Satorra-Saris法或Monte Carlo法进行先验检验力分析时，应当审慎确定参数的大小，必要时，在缺乏既往研究或理论的支持下，研究者可以反复设定参数的值，并从结果中选取一个较保守的样本量。

5.2 理论协方差矩阵的产生方式

如前所述，结构方程模型统计检验力分析的Monte Carlo法和Satorra-Saris法在分析时需要指定理论协方差矩阵，且理论协方差矩阵的指定存在两种不同模式：直接指定或通过指定备择模型生成。虽然从分析过程和分析结果上二者可能并无不同之处，但从逻辑上讲，不同的理论协方差矩阵生成方式意味着对结构方程建模过程中误差来源的不同理解。Jak等人(2021)指出作为衡量理论模型与真实数据间差异的似然比统计量 T_{ML} 实际上包含了两部分的误差：随机抽样造成的抽样误差和理论模型与真实数据不匹配的误差。事实上，理论模型与真实数据不匹配的误差还可以再分为两种不同的来源：来自模型内部参数设定错误以及模型本身对总体的简化。当研究者通过指定备择模型的方式生成理论协方差矩阵时，研究者实际上默认了模型内部参数设定错误是导致误差的主要来源，而当研究者通过直接指定的方式生成理论协方差矩阵时，则将这两种不同来源的误差“打包”在了一起。事实上，经典的Satorra-Saris法与MacCallum法之间也具有类似的关系，经典的Satorra-Saris法更加强调来自已经结构化了的模型内部的误差，MacCallum法则通过模型拟合指数将模型内部的误差和模型对总体简化的误差放在一起考察。这种细微的差别意味着研究者需要在面对不同检验的情况下灵活选择理论协方差矩阵的生成方式：当有较强的理论或既往研究指向误差可能存在于模型内部时，使用指定备择模型的生成方式；而当研究者对这些信息掌握程度较少时，使用直接指定的方式或干脆换用MacCallum法进行分析。

此外，另一个有趣的议题是，使用先前的研究信息作为指定理论协方差的依据是否是合适的。诚然，既往研究中确实存有使用相似主题的研究结果作为统计检验力分析过程中设定参数依止的先例，但有研究者指出，这种指定方式可能错误地在统计检验力分析过程中混入了抽样误差而降低了统计检验力的可靠性，并进一步提出了纠正该误差的方案(Liu & Wang, 2019)。对该问题的深入讨论已超出了本文的范畴，有兴趣的读者可以进一步阅读Liu和Wang (2019)的研究。

5.3 非中心 χ^2 分布近似带来的误差

Satorra-Saris法与MacCallum法这两种分析方法从本质上来说都是利用非中心 χ^2 分布作为当 H_1 成立时样本似然比统计量 T_{ML} 分布的近似。Steiger等人(1985)指出，这种近似当且仅当样本满足似然函数关于分布的基本假设、样本量不太小、拟合函数值不是太大的情况下成立。诚然，Satorra等人(1991)的研究进一步指出，当样本服从正态分布且理论模型与实际数据的偏离程度不是十分严重时，即使对小样本而言，这种利用非中心 χ^2 分布进行近似的精度也是基本令人满意的。但对于分布假设不满足时这种近似的精度是否仍令人满意，目前仍缺乏相关研究的分析。这意味着研究者在对结构方程模型的统计检验力进行分析时，应当充分了解样本分布情况与所选结构方程模型参数估计方法之间的适配性，在进行统计检验力分析和实际进行数据处理时使用同样的估计方法，在样本分布偏离正态分布的情况下，也可以考虑换用适用范围更广的Monte Carlo法进行统计检验力分析。目前主流软件提供的参数估计方法中，极大似然法(maximum likelihood, ML)和广义最小二乘法(generalized least squares, GLS)要求数据为正态分布，加权最小二乘法(weighted least squares, WLS)、非加权最小二乘法(unweighted least squares, ULS)和对角线加权最小二乘法(diagonally weighted least squares, DWLS)对数据分布要求不高，但ULS属于尺度依赖方法(scale-dependent method)，使用时最好保证各个变量的测量单位统一，而WLS和DWLS通常则需要1000以上的样本量(吴明隆, 2010)。

此外，Satorra-Saris法在使用非中心 χ^2 分布对样本似然比统计量 T_{ML} 的分布进行近似时，同时使用模型拟合“真实总体”得到的统计量作为对非中心参数的估计。这意味着使用Satorra-Saris法分析模型参数检验的统计检验力时，检验方法本身也会对结果产生影响。本文4.2部分模型参数检验的先验检验力分析中，Monte Carlo法与Satorra-Saris法对同一检验的分析得出了差异较

ChinaXiv合作期刊

chinaXiv:202303.09766v1

大的结果就是这个原因。在该例中, 在设定 α = 0.05, N = 401 的情况下, 使用 Wald 检验、LM 检验和似然比检验得到的统计量对非中心参数进行估计时得到的非中心参数分别为 13.399, 22.658 和 23.265, 对应的统计检验力分别为 0.955, 0.997 和 0.998。使用 Monte Carlo 法对三种方法进行统计检验力分析, 得到的结果则分别为 0.996, 0.998 和 0.998。这提示研究者在使用 Satorra-Saris 法分析 Wald 检验的统计检验力时, 可能会由于低估统计检验力而高估所需要的样本数量。

5.4　Monte Carlo 法的精度与计算效率

诚然 Monte Carlo 法由于可以广泛适用于各类检验场景, 且在模拟次数足够多的情况下得出的统计检验力的估计值依概率收敛于其理论真值, 使得该方法广泛受到研究者的青睐。但 Monte Carlo 法同样面临着结果不稳定以及计算效率低下的情形。与其他两种方法不同, Monte Carlo 法得到的计算结果并不是一个确定值, 而是一个随机变量, 该随机变量的期望值等于统计检验力的真实值, 这就导致了对同一个假定样本量进行多次统计检验力分析时可能会得出不同的结果。这意味着研究者在使用 Monte Carlo 法进行统计检验力分析时求得的样本量理应为一约数而非确切值。换言之, Monte Carlo 法无法像 Satorra-Saris 法或 MacCallum 法那样给出“理论上达到目标统计检验力所需的最小样本量”。这也同时意味着通过缩短采样间距的方式对基于 Monte Carlo 法的统计检验力分析的精度的提升是有限的。

此外, Monte Carlo 法的计算效率一直以来也是比较困扰研究者的一个问题。与 Satorra-Saris 法和 MacCallum 法误差不同, Monte Carlo 法的误差来源主要是模拟样本带来的偶然误差。根据 De Moivre-Laplace 中心极限定理, 随着重复次数 R 的提升, Monte Carlo 法所得到的统计检验力估计值的标准差会降低, 但其降低速率与重复次数 R 之间并非线性关系, 而是二次关系。换言之, 若研究者希望将 Monte Carlo 法得到的估计值的精度提高一个数量级(相当于让估计值的标准差降低为原来的 0.1 倍), 则需要将 R 扩大为原来的 100 倍。这大大降低了 Monte Carlo 法的计算效率。并且, Monte Carlo 法本身计算起来就比较耗时, 在 CPU 主频为 2.50 GHz 的笔记本电脑上运行本文中关于中介模型统计检验力分析的 R 程序, 当设定重复次数 R = 1000, 样本量 N = 401 时, 使用 Monte Carlo 法分析 Wald 检验统计检验力所需的时间大致为 60 s (单线程); 若将检验方法更换为 LM 检验或似然比检验, 执行同样的分析所需时间则上升到约 900 s (单线程); 若对非参数 Bootstrap 检验的统计检验力进行分析, 即使采用并行计算技术将任务分配到多线程计算机的不同线程上处理, 所需时间通常也会超过一个小时。此时若反复设定样本量甚至参数进行多轮分析从时间上来说是不经济的。诚然有学者指出此时可以换用参数 Bootstrap 法(或与之等价的 MC 法)简化计算过程, 节省计算时间(Schoemann et al., 2017), 但这种更换也将在一定程度上造成统计检验力估计时的误差。

5.5　不同拟合指数的选择对 MacCallum 法的影响

本文主要介绍了用 MacCallum 法对 RMSEA 相关的零假设进行统计检验力分析的理论与实践。事实上, MacCallum 法可以用于多种拟合指数的情形。MacCallum 和 Hong (1997)将该方法进一步扩展到了针对 GFI 和 AGFI 相关零假设的情形。Kim (2005)则提出了当零假设是关于 RMSEA、CFI、MFI 和 GFI 这四种参数其一时估算样本量的简便公式。值得注意的是, 上述参数间的换算关系还会受到模型本身的影响。因此, 目前学界主流建议的拟合指数界值在不同的模型中并不等价, 这就导致了针对同一个计划模型, 采用不同拟合指数进行先验检验力分析时, 可能得出不同的结果。例如, 根据温忠麟等人(2004)的建议, 评估模型拟合时, MFI 的建议界值为 0.85, RMSEA 的建议界值为 0.08。此时, 对于不同自由度的模型而言, 依据这两个指标的建议界值计算出的计划样本量也有所不同。以计划显著性水平为 0.05, 统计检验力为 0.95 的情况为例, 当模型自由度为 10 时, 根据 MFI 和 RMSEA 计算出的计划样本分别有 77 人和 383 人; 当模型自由度为 50 时, 根据 MFI 和 RMSEA 计算出的计划样本分别有 134 人和 136 人; 而当模型自由度为 100 时, 根据 MFI 和 RMSEA 计算出的计划样本则分别有 176 人和 90 人。因此, 本文建议研究者在使用 MacCallum 法进行先验检验力分析时, 使用多种拟合指数构建假设进行检验, 并在所有计划样本量中选取最大值。至于拟合指数的具体选择, 可以参考温忠麟等人(2004)的研究。

ChinaXiv合作期刊

5.6 评价模型拟合时的统计检验力分析与等效性检验

如前所述，既往研究中对传统 χ^2 检验的一个批评就是该检验将想要验证的假设作为零假设，在结果报告中又缺乏对其Ⅱ类错误发生概率的分析，使其无法真正起到验证模型的作用(王阳等, 2020)。而等效性检验实质上就是通过构建一个“可接受的差异”标准的方式重写零假设，从而将想要验证的假设重新放回了备择假设的位置，在一定程度上，也可以将等效性检验视作一种“颠倒了 H_0 与 H_1 的传统 χ^2 检验”。根据 MacCallum 等人(1996)的研究，可以推知，对 H_{0a}: $F_{ML0} \leqslant k_0$，且真实参数为 $F_{ML0} = k_A$ 的情形，存在一种与之“对偶”的情形: H_{0b}: $F_{ML0} \geqslant k_A$，且此时的真实参数为 $F_{ML0} = k_0$，这两种情形下Ⅰ类错误概率和Ⅱ类错误概率互相为对方的Ⅱ类错误概率与Ⅰ类错误概率。这意味着，对传统 χ^2 检验而言，在进行先验检验力分析时，若研究者参考等效性检验中的 H_0 设定 F_{ML} 的真值(即 F_{ML} 的真值不小于 ε_0)，并设定显著性水平为 α，统计检验力为 1−β 计算出计划样本量 N。那么 N 同样可以保证在控制显著性水平为 β 的水平下，对实际拟合良好的模型(即 $F_{ML0} = 0$)，采用零假设为 H_0: $F_{ML0} \geqslant \varepsilon_0$ 的等效性检验进行检验时，正确拒绝零假设的概率为 1−α。因此，本文建议研究者在对评估模型拟合相关的假设检验进行统计检验力分析时，设定不小于 0.95 的统计检验力。这样可以保障在模型拟合较差的情况下，研究者错误地认为模型拟合良好的可能性不高于 0.05。

6 总结

本文主要介绍了结构方程模型假设检验统计检验力分析的三种代表方法：Satorra-Saris 法、MacCallum 法和 Monte Carlo 法，这些方法各有自己的适用情形。相比于目前已被广大研究者所熟知的平均数差异检验的统计检验力分析，结构方程模型的统计检验力分析对研究者的统计水平有更高的要求，并且检验过程较为复杂。综合考虑易掌握和易使用两方面的要求，本文建议研究者在研究设计的过程中根据自己需要进行的检验类型选择恰当的统计检验力分析方式，具体可以参考图 2 进行选择。

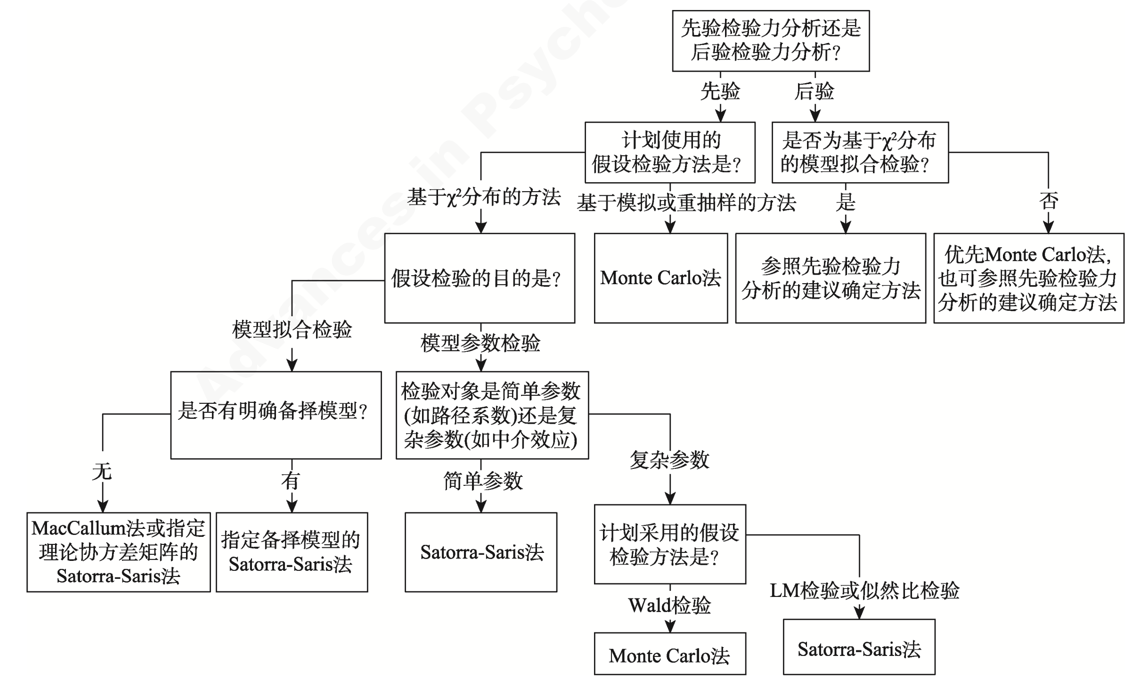


图 2 结构方程模型统计检验力分析方法选择建议

注：未标注理论协方差矩阵生成方法的表示研究者可以参照 5.2 部分的建议自行设定

chinaXiv:202303.09766v1

ChinaXiv合作期刊

致谢：感谢刘力沛博士对本文英文摘要的修改，感谢崔虞馨、姜鹤、荣杨在本文修改、定稿过程中提供的诸多帮助，感谢两位匿名审稿人对本文提供的宝贵意见。

参考文献

方杰，张敏强，李晓鹏. (2011). 中介效应的三类区间估计方法. *心理科学进展, 19*(5), 765−774.

屠金路，金瑜，王庭照. (2005). Bootstrap 法在合成分数信度区间估计中的应用. *心理科学, 28*(5), 1199−1200.

王阳，温忠麟，付媛姝. (2020). 等效性检验——结构方程模型评价和测量不变性分析的新视角. *心理科学进展, 28*(11), 1961−1969.

文书锋，汤冬玲，俞国良. (2009). 情绪调节自我效能感的应用研究. *心理科学, 32*(3), 666−668.

温忠麟，侯杰泰，马什赫伯特. (2004). 结构方程模型检验: 拟合指数与卡方准则. *心理学报, 36*(2), 186−194.

温忠麟，叶宝娟. (2014a). 有调节的中介模型检验方法:竞争还是替补? *心理学报, 46*(5), 714−726.

温忠麟，叶宝娟. (2014b). 中介效应分析: 方法和模型发展. *心理科学进展, 22*(5), 731−745.

吴明隆. (2010). *结构方程模型——AMOS 的操作与应用* (第 2 版). 重庆大学出版社.

吴艳，温忠麟. (2011). 与零假设检验有关的统计分析流程. *心理科学, 34*(1), 230−234.

叶宝娟，温忠麟. (2012). 用 Delta 法估计多维测验合成信度的置信区间. *心理科学, 35*(05), 1213−1217.

张涵，康飞. (2016). 基于 Bootstrap 的多重中介效应分析方法. *统计与决策, 32*(5), 75−78.

张莉，薛香娟，赵景欣. (2019). 歧视知觉、抑郁和农村留守儿童的学业成绩：纵向中介模型. *心理科学, 42*(3), 584−590.

赵礼，王晖. (2019). 统计检验力的分析流程与多层模型示例. *心理技术与应用, 7*(5), 276−283.

Chou, C.-P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, lagrange multiplier, and wald tests. *Multivariate Behavioral Research, 25*(1), 115−136.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics, 7*(1), 1−26.

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods, 6*(3), 258−269.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician, 55*(1), 19−24.

Hollerbach, P., Johansson, A., Ventus, D., Jern, P., Neumann, C. S., Westberg, L., ... Mokros, A. (2018). Main and interaction effects of childhood trauma and the MAOA uVNTR polymorphism on psychopathy. *Psychoneuroendocrinology, 95*, 106−112.

Jak, S., Jorgensen, T. D., Verdam, M., Oort, F. J., & Elffers, L. (2021). Analytical power calculations for structural equation modeling: A tutorial and shiny app. *Behavior Research Methods, 53*(4), 1385−1406.

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 12*(3), 368−390.

Kornadt, A. E., Siebert, J. S., & Wahl, H.-W. (2019). The interplay of personality and attitudes toward own aging across two decades of later life. *PLoS ONE, 14*(10), Article e0223622. https://doi.org/10.1371/journal.pone.0223622

Liu, X., & Wang, L. J. (2019). Sample size planning for detecting mediation effects: A power analysis procedure considering uncertainty in effect size estimates. *Multivariate Behavioral Research, 54*(6), 822−839.

MacCallum, R. C., Browne, M. W., & Li, C. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods, 11*(1), 19−35.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130−149.

MacCallum, R. C., & Hong, S. (1997). Power analysis in covariance structure modeling using GFI and AGFI. *Multivariate Behavioral Research, 32*(2), 193−210.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99−128.

Marcoulides, K. M., & Yuan, K.-H. (2017). New ways to evaluate goodness of fit: A note on using equivalence testing to assess structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(1), 148−153.

Moshagen, M., Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 54−60.

Mullen, P. R., & Crowe, A. (2017). Self-stigma of mental illness and help seeking among school counselors. *Journal of Counseling & Development, 95*(4), 401−411.

Muthén, B. O. (2004). *Mplus technical appendices*. CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal,*

chinaXiv:202303.09766v1

ChinaXiv合作期刊

9(4), 599−620.
Preacher, K. J., & Coffman, D. L. (2006). *Computing power and minimum sample size for RMSEA* [Computer software]. Retrieved September 10, 2021, from http://quantpsy.org/rmsea/rmsea.htm
Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures, 6*(2), 77−98.
Rosseel, Y. (2021). *The lavaan tutorial*. Retrieved September 10, 2021, from https://lavaan.ugent.be/tutorial/tutorial.pdf
Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*(1), 131−151.
Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*(1), 83−90.
Satorra, A., Saris, W. E., & Pijper, W. M. (1991). A comparison of several approximations to the power function of the likelihood ratio test in covariance structure analysis. *Statistica Neerlandica, 45*(2), 173−185.
Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science, 8*(4), 379−386.
Steiger, J., Shapiro, A., & Browne, M. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50*(3), 253−263.
Venables, W. (1975). Calculation of confidence intervals for noncentrality parameters. *Journal of the Royal Statistical Society: Series B, 37*(3), 406−412.
Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science, 4*(1), 1−17.
Yuan, K.-H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests. *Psychological Methods, 21*(3), 405−426.
Yuan, K.-H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(3), 319−330.
Zhai, H.-K., Li, Q., Hu, Y.-X., Cui, Y.-X., Wei, X.-W., & Zhou, X. (2021). Emotional creativity improves posttraumatic growth and mental health during the COVID-19 pandemic. *Frontiers in Psychology, 12*, Article 600798.https://doi.org/10.3389/fpsyg.2021.600798
Zhang, X. J., & Savalei, V. (2016). Bootstrapping confidence intervals for fit indexes in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(3), 392−408.
Zhang, X., & Zaman, B. U. (2020). Adoption mechanism of telemedicine in underdeveloped country. *Health Informatics Journal, 26*(2), 1088−1103.
Zhang, Z. Y. (2014). Monte Carlo based statistical power analysis for mediation models: Methods and software. *Behavior Research Methods, 46*(4), 1184−1198.
Zhang, Z. Y., & Wang, L. J. (2020). *bmem: Mediation analysis with missing data using bootstrap* [Computer software]. Retrieved December 11, 2021, from https://CRAN.R-project.org/package=bmem

Power analysis in structural equation modeling: Principles and methods

ZHAI Hongkun, LI Qiang, WEI Xiaowei
(*Department of Social Psychology, Nankai University, Tianjin* 300350, *China*)

Abstract: Structural equation modeling (SEM) is an important statistical tool in psychology, management, and sociology. However, the lack of analyses and reports of statistical power in many studies using SEM has reduced the probative power of the study results. There are three main types of statistical power analysis methods for SEM: the Satorra-Saris, MacCallum, and Monte Carlo methods. The Satorra-Saris method applies to cases where the alternative model is clear, analysis object is simple, and the statistic to be examined is based on the χ^2 distribution. The MacCallum method is applicable to the case of χ^2-based overall model fit tests with unknown alternative models. Further, the Monte Carlo method applies to cases where the analysis object is complex, or the test is performed using simulation or resampling methods. In practice, researchers should clarify the purposes and methods of the test as well as the availability of alternative models as a priority. The research method can then be determined on the basis of the above information.

Key words: structural equation model, statistical power, fitting test, test of model parameters

chinaXiv:202303.09766v1

ChinaXiv合作期刊

《结构方程模型统计检验力分析：原理与方法》补充材料

◎补充材料 A：案例部分的 R 代码

```
#载入程序包
if(!require(lavaan)){install.packages('lavaan')}
if(!require(mvtnorm)){install.packages('mvtnorm')}
if(!require(semPower)){install.packages('semPower')}
#==============================================================================
#4.1 模型拟合检验的先验检验力分析
#4.1.1 Satorra-Saris 法
#(1) 设定用于生成理论协方差矩阵的模型
genMod = '
POS ~~ 0.23 * DES
POS ~~ 0.40 * ANG
DES ~~ 0.78 * ANG
POS =~ 0.87 * X1 + 0.90 * X2 + 0.75 * X3 + 0.55 * X4
DES =~ 0.59 * X5 + 0.72 * X6 + 0.58 * X7 + 0.71 * X8 + 0.59 * X9
ANG =~ 0.73 * X10 + 0.78 * X11 + 0.57 * X12
X1 ~~ 0.24 * X1
X2 ~~ 0.19 * X2
X3 ~~ 0.44 * X3
X4 ~~ 0.70 * X4
X5 ~~ 0.65 * X5
X6 ~~ 0.47 * X6
X7 ~~ 0.54 * X7
X8 ~~ 0.49 * X8
X9 ~~ 0.52 * X9
X10 ~~ 0.45 * X10
X11 ~~ 0.39 * X11
X12 ~~ 0.55 * X12'
#(2) 设定用于分析的模型
anaMod = '
POS =~ X1 + X2 + X3 + X4
DES =~ X5 + X6 + X7 + X8 + X9
ANG =~ X10 + X11 + X12'
#(3) 生成理论协方差矩阵
E1 = cfa(genMod)@implied$cov[[1]]
rownames(E1) = paste0('X', 1:12)
#(4) 分析模型
fit = cfa(anaMod, sample.cov = E1, sample.nobs = 101, likelihood = 'Wishart')
#(5) 提取关键参数
f0 = fit@optim$fx * 2
#(6) 用于计算等效性检验样本量的函数
```

chinaXiv:202303.09766v1

Advances in Psychological Science

ChinaXiv合作期刊


```
setPowerFindEqTestN = function(f0, df, alpha = 0.05, beta = 0.8, e = 0.08){
    foo = function(N){pchisq(qchisq(alpha, df, ncp = e*e*df*N), df, f0*N) - beta}
    n = uniroot(foo, c(10, 1e4))$root
    return(ceiling(n+1))
}
#(7) 执行计算
setPowerFindEqTestN(f0, df = 51, alpha = 0.05, beta = 0.95)
#-------------------------------------------------------------------------------
#4.1.2 MacCallum 法
semPower.aPriori(effect = 0.08, effect.measure = 'RMSEA', alpha = 0.05,
                power = 0.95, df = 51) -> RES
summary(RES)
#-------------------------------------------------------------------------------
#4.1.3 Monte Carlo 法
#(1) 生成理论协方差矩阵的部分与 Satorra-Saris 法相同, 这里直接使用了前面的结果
#(2) 用于计算统计检验力的程序
mcCFA = function(N, E, mod, R = 1000L){
    RES = matrix(0, nrow = R, ncol = 62)
    for(i in 1:R){
        Sample = rmvnorm(N, sigma = E)
        if(!is.null(colnames(E))){
            colnames(Sample) = colnames(E)
        }else{
            colnames(Sample) = rownames(E)
        }
        RES[i, ] = tryCatch({
            tempFit = cfa(mod, Sample, std.lv = TRUE, likelihood = 'wishart')
            ind = sum(unlist(parTable(tempFit)[, 14:15]))
            ind = ind / ind
            c(!is.nan(ind), unlist(parTable(tempFit)[, 14:15]),
               as.numeric(fitmeasures(tempFit, 'chisq')))
        }
        , warning = function(w){c(0, rep(NaN, 61))})
    }
    return(RES)
}
#(3) 利用程序计算统计检验力
set.seed(1234)
mcCFA(101, E1, anaMod) -> RES1
mcCFA(201, E1, anaMod) -> RES2
#(4) 统计结果, 并进行规范化输出
resCFA = function(RES, TV, N){
    R = nrow(RES)
```

chinaXiv:202303.09766v1

ChinaXiv合作期刊


```
    P = sum(RES[, 1]) / R
    RES = RES[which(RES[, 1] == 1), -1]
    COV = matrix(0, nrow = nrow(RES), ncol = 30)
    for(i in 1:nrow(RES)){for(j in 1:30){COV[i, j] = (TV[j] <= qnorm(0.975, RES[i, j],
        RES[i, j+30])) & (TV[j] >= qnorm(0.025, RES[i, j], RES[i, j+30]))}}
    ME1 = apply(RES[, 1:30], 2, mean)
    ME2 = apply(RES[, 31:60], 2, mean)
    SE = apply(RES[, 1:30], 2, sd)
    PC = apply(COV, 2, mean)
    PM = cbind(ME1, ME2, SE, PC, abs(ME1 - TV) / TV, abs(ME2 - SE) / SE)
    RES = RES[, 61]
    Power = sum(RES <= qchisq(0.05, 51, 0.08*0.08*51*(N-1))) / length(RES)
    colnames(PM) = c('MC_MEAN', 'MC_SE', 'MEAN_SE', 'Coverage', 'Mean_Bias',
                     'SE_Bias')
    RES = list(Proper = P, CHECK = PM, POWER = Power)
    return(RES)
}
#(5) 给出参数的真实值
TV = parTable(cfa(anaMod, sample.cov = E1, sample.nobs = 101,
     std.lv = TRUE, likelihood = 'Wishart'))[, 14]
#(6) 给出规范化输出
resCFA(RES1, TV, 101)
resCFA(RES2, TV, 201)
#(7) 补充中点采样
mcCFA(151, E1, anaMod) -> RES3
resCFA(RES3, TV, 151)
#-------------------------------------------------------------------------------
#4.2 模型参数检验的先验检验力分析
#4.2.1 Satorra-Saris 法; Wald 检验
#(1) 设定理论协方差矩阵
E2 = matrix(c(
1.0, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1,
0.1, 1.0, 0.1, 0.1, 0.1, 0.1, 0.1,
0.1, 0.1, 1.0, 0.1, 0.1, 0.1, 0.1,
0.1, 0.1, 0.1, 1.0, 0.3, 0.3, 0.3,
0.1, 0.1, 0.1, 0.3, 1.0, 0.3, 0.3,
0.1, 0.1, 0.1, 0.3, 0.3, 1.0, 0.3,
0.1, 0.1, 0.1, 0.3, 0.3, 0.3, 1.0
), nrow = 7)
colnames(E2) = c('Age', 'Ser', 'STR', 'ECI', 'PSS', 'RES', 'SRG')
#(2) 设定分析模型
anaMod = '
SRG ~ b*RES + d*PSS + ECI + STR + Ser + Age
```

chinaXiv:202303.09766v1

ChinaXiv合作期刊


```
RES ~ k*PSS + ECI + STR + Ser + Age
PSS ~ a*ECI + e * STR + Ser + Age
akb := a*k*b'
#(3) 用于计算统计检验力的程序
ssSEM = function(E, mod, PID, alpha = 0.05, beta = 0.95){
    foo = function(N){
        fit = sem(mod, sample.cov = E, sample.nobs = N, likelihood = 'Wishart')
        NCP = (parTable(fit)[PID, 14] / parTable(fit)[PID, 15]) ** 2
        return(pchisq(qchisq(1 - alpha, 1), 1, NCP, FALSE) - beta)
    }
    n = uniroot(foo, c(10, 1e4))$root
    return(ceiling(n + 1))
}
#(4) 执行计算
ssSEM(E2, anaMod, 29)
#-----------------------------------------------------------------------------
#4.2.2 Monte Carlo 法；Wald 检验
#(1) 生成理论协方差矩阵的部分与 Satorra-Saris 法相同，这里直接使用了前面的结果
#(2) 用于计算统计检验力的程序
mcSEM = function(N, E, mod, R = 1000L){
#注：基于无完全共线性的真实数据协方差矩阵的显变量回归理论上不可能出现不恰当解，因此没有考虑对其进行统计
    RES = matrix(0, nrow = R, ncol = 58)
    for(i in 1:R){
        Sample = rmvnorm(N, sigma = E)
        if(!is.null(colnames(E))){
            colnames(Sample) = colnames(E)
        }else{
            colnames(Sample) = rownames(E)
        }
        for(j in 1:ncol(Sample)){Sample[, j] = (Sample[, j] - mean(Sample[, j])) /
           sd(Sample[, j])}
        tempFit <- sem(mod, Sample, likelihood = 'wishart')
        RES[i, ] = unlist(parTable(tempFit)[, 14:15])
    }
    return(RES)
}
#(3) 利用程序计算统计检验力
set.seed(4321)
mcSEM(601, E2, anaMod) -> RES1
mcSEM(501, E2, anaMod) -> RES2
mcSEM(401, E2, anaMod) -> RES3
mcSEM(301, E2, anaMod) -> RES4
```

chinaXiv:202303.09766v1

ChinaXiv合作期刊


```
#(4) 统计结果，并进行规范化输出
resSEM = function(RES, TV){
    R = nrow(RES)
    COV = matrix(0, nrow = nrow(RES), ncol = 29)
    for(i in 1:nrow(RES)){for(j in 1:29){COV[i, j] = (TV[j] <= qnorm(0.975, RES[i, j],
        RES[i, j+29])) & (TV[j] >= qnorm(0.025, RES[i, j], RES[i, j+29]))}}
    ME1 = apply(RES[, 1:29], 2, mean)
    ME2 = apply(RES[, 30:58], 2, mean)
    SE = apply(RES[, 1:29], 2, sd)
    PC = apply(COV, 2, mean)
    PM = cbind(ME1, ME2, SE, PC, abs(ME1 - TV) / TV, abs(ME2 - SE) / SE)
    Power = sum(abs(RES[, 29] / RES[, 58]) >= qnorm(0.975)) / R
    colnames(PM) = c('MC_MEAN', 'MC_SE', 'MEAN_SE', 'Coverage', 'Mean_Bias',
                    'SE_Bias')
    RES = list(CHECK = PM[1:15, ], POWER = Power)
    return(RES)
}
#(5) 给出参数的真实值
TV = parTable(cfa(anaMod, sample.cov = E2, sample.nobs = 101,
     likelihood = 'Wishart'))[, 14]
#(6) 给出规范化输出
resSEM(RES1, TV)
resSEM(RES2, TV)
resSEM(RES3, TV)
resSEM(RES4, TV)
# 结果显示 N = 301/601 时精度不足
#(7) 重新抽取精度不足的样本
mcSEM(601, E2, anaMod) -> RES1
mcSEM(301, E2, anaMod) -> RES4
resSEM(RES1, TV)
resSEM(RES4, TV)
```

◎补充材料 B：统计检验力的后验分析实例

本部分的演示样例同样来自 Zhai 等人(2021)的研究。该研究样本量为 439 人，其中男 157 人，女 282 人。本部分同样以该研究中对情绪调节自我效能感量表的验证性因子分析与以压力后成长为因变量的链式中介模型为分析对象，对其进行后验统计检验力分析。该研究中，关于情绪调节自我效能感量表的验证性因子分析结果为: $\chi^2 = 151.979$, $df = 51$, RMSEA = 0.067，参考王阳等人(2020)的建议，以 $\text{RMSEA}_0 = 0.08$ 为经验临界值进行等效性检验得 $p = 0.046$。链式中介模型所涉及变量间的协方差矩阵参见表 B-1，具体的模型设置同正文部分，路径“情绪创造力→领悟社会支持→情绪调节自我效能感→压力后成长”中介效应值为 0.009(标准化结果为 0.021), Wald 检验 $z = 2.713$, $p = 0.007$。参考正文图 2 中的建议，综合确定使用 MacCallum 法对模型拟合检验部分进行后验检验力分析，使用 Monte Carlo 法对链式中介效应部分进行后验检验力分析。

chinaXiv:202303.09766v1

ChinaXiv合作期刊

表 B-1. 链式中介模型所涉及变量间的协方差矩阵

	1	2	3	4	5	6	7
1. 年龄	36.818						
2. 疫情严重程度	−0.194	0.936					
3. 疫情相关压力	−0.643	2.788	215.412				
4. 情绪创造力	−9.107	0.098	9.321	129.108			
5. 领悟社会支持	−3.390	−0.992	0.305	60.613	100.388		
6. 情绪调节自我效能感	−1.217	−0.294	2.198	42.998	29.468	31.922	
7. 压力后成长	−1.981	−0.275	−6.180	24.855	21.677	11.936	21.294

1. 模型拟合检验的后验检验力分析

在后验检验力分析中，通常直接以得出的结果作为真值。故本例相当于检验样本量为 439、$\alpha = 0.05$、RMSEA 真值为 0.067 时，样本的 χ^2 统计量落在等效性检验零假设 H_0: $\text{RMSEA}_0 \geq 0.08$ 拒绝域中的概率。根据正文中的公式(1a)和(2a)，可以求出后验检验力，对应的 R 脚本如下：

```
ncp0 = (439-1) * 51 * 0.080**2
ncpa = (439-1) * 51 * 0.067**2
Tmlc = qchisq(0.05, 51, ncp0)
(POW = pchisq(Tmlc, 51, ncpa))
```

运行上述代码，可得后验检验力为 0.546。

2. 中介效应检验的后验检验力分析

以对“情绪创造力→领悟社会支持→情绪调节自我效能感→压力后成长”这一路径中介效应的 Wald 检验为分析对象，控制 I 类错误水平 α 为 0.05，设定 *N* 为 439 进行分析，得到的结果全部为恰当解，对应的统计检验力为 0.847，对应的各参数估计情况、相对误差以及 95%置信区间覆盖率参见表 B-2。本例中关键参数的估计误差均不超过 5%，95%置信区间覆盖率均处于 0.91 至 0.98 之间，部分参数的均值误差较大，这主要是由于这些参数的理论值过小导致的，实际上这些参数标准误的误差均不超过 5%，并且这些参数都是与控制变量相关的参数，说明 Monte Carlo 模拟结果仍具可信度。这部分的 R 脚本为：

```
#载入程序包
if(!require(lavaan)){install.packages('lavaan')}
#读入样本协方差矩阵
E = matrix(c(
36.818, -0.194,  -0.643,  -9.107,  -3.390, -1.217, -1.981,
-0.194,  0.936,   2.788,   0.098,  -0.992, -0.294, -0.275,
-0.643,  2.788, 215.412,   9.321,   0.305,  2.198, -6.180,
-9.107,  0.098,   9.321, 129.108,  60.613, 42.998, 24.855,
-3.390, -0.992,   0.305,  60.613, 100.388, 29.468, 21.677,
-1.217, -0.294,   2.198,  42.998,  29.468, 31.922, 11.936,
-1.981, -0.275,  -6.180,  24.855,  21.677, 11.936, 21.294), nrow = 7)
rownames(E) = c('Age', 'Ser', 'STR', 'ECI', 'PSS', 'RES', 'SRG')
#设定分析模型
anaMod = '
SRG ~ b*RES + d*PSS + ECI + STR + Ser + Age
RES ~ k*PSS + ECI + STR + Ser + Age
PSS ~ a*ECI + e * STR + Ser + Age
```

chinaXiv:202303.09766v1

ChinaXiv合作期刊


```
akb := a*k*b'
TV = parTable(sem(anaMod, sample.cov = E, sample.nobs = 439, likelihood = 'Wishart'))[, 14]
#执行分析, 这里利用wishart分布替代了原始数据
mcSEM = function(N, E, mod, R = 1000L){
    RES = matrix(0, nrow = R, ncol = 58)
    SaS = rWishart(R, N-1, E) / (N-1)
    for(i in 1:R){
        tempE = SaS[, , i]
        rownames(tempE) = rownames(E)
        tempFit = sem(mod, sample.nobs = N, sample.cov = tempE, likelihood = 'wishart')
        RES[i, ] = unlist(parTable(tempFit)[, 14:15])}
    return(RES)}
set.seed(1234)
RES = mcSEM(439, E, anaMod)
resSEM = function(RES, TV){
    R = nrow(RES)
    COV = matrix(0, nrow = nrow(RES), ncol = 29)
    for(i in 1:nrow(RES)){for(j in 1:29){
        COV[i, j] = (TV[j] <= qnorm(0.975, RES[i, j], RES[i, j+29])) &
         (TV[j] >= qnorm(0.025, RES[i, j], RES[i, j+29]))}}
    ME1 = apply(RES[, 1:29], 2, mean)
    ME2 = apply(RES[, 30:58], 2, mean)
    SE = apply(RES[, 1:29], 2, sd)
    PC = apply(COV, 2, mean)
    PM = cbind(ME1, ME2, SE, PC, abs(ME1 - TV) / TV, abs(ME2 - SE) / SE)
    Power = sum(abs(RES[, 29] / RES[, 58]) >= qnorm(0.975)) / R
    colnames(PM) = c('MC_MEAN', 'MC_SE', 'MEAN_SE', 'Coverage', 'Mean_Bias',
                     'SE_Bias')
    RES = list(CHECK = PM[1:15, ], POWER = Power)
    return(RES)}
resSEM(RES, TV)
```

表 B-2.　中介效应 Monte Carlo 检验的参数估计情况

路径	理论值	估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差
RES→SRG	**0.145**	**0.144**	**0.045**	**0.045**	**0.949**	**0.002**	**0.002**
PSS→SRG	0.118	0.118	0.022	0.022	0.953	0.002	0.023
ECI→SRG	0.090	0.090	0.023	0.023	0.945	0.003	0.006
STR→SRG	−0.034	−0.033	0.013	0.012	0.958	−0.012	0.013
Ser→SRG	−0.035	−0.047	0.193	0.195	0.949	−0.324	0.009
Age→SRG	−0.017	−0.018	0.030	0.031	0.945	−0.065	0.023
PSS→RES	**0.126**	**0.127**	**0.023**	**0.022**	**0.961**	**0.009**	**0.035**
ECI→RES	0.277	0.278	0.020	0.021	0.941	0.002	0.014
STR→RES	0.001	0.001	0.013	0.014	0.943	0.574	0.010

chinaXiv:202303.09766v1

ChinaXiv合作期刊

续表

路径	理论值	估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差
Ser→RES	−0.203	−0.201	0.205	0.201	0.950	−0.007	0.018
Age→RES	0.046	0.046	0.032	0.032	0.949	0.007	0.011
ECI→PSS	**0.472**	**0.470**	**0.036**	**0.036**	**0.944**	**0.004**	**0.008**
STR→PSS	−0.005	−0.004	0.028	0.029	0.942	−0.126	0.021
Ser→PSS	−1.091	−1.091	0.423	0.431	0.941	0.000	0.017
Age→PSS	0.019	0.023	0.067	0.068	0.946	0.198	0.021

注：加粗的表示该路径系数是关键参数；RES：情绪调节自我效能感，PSS：领悟社会支持，ECI：情绪创造力，SRG：压力后成长，STR：疫情相关压力，Ser：疫情严重程度，Age：年龄

参考文献

Zhai, H.-K., Li, Q., Hu, Y.-X., Cui, Y.-X., Wei, X.-W., & Zhou, X. (2021). Emotional creativity improves posttraumatic growth and mental health during the COVID-19 pandemic. *Frontiers in Psychology, 12*, Article 600798. https://doi.org/10.3389/fpsyg.2021.600798

王阳，温忠麟，付媛姝. (2020). 等效性检验——结构方程模型评价和测量不变性分析的新视角. *心理科学进展, 28*(11), 1961−1969.

◎**补充材料 C: MacCallum 法适用的拟合指数及其与非中心参数的换算公式**

根据正文部分的论述，不难发现 MacCallum 法的核心在于将拟合指数根据其定义式逆变换为非中心参数，作为对 δ 的估计值。同时，根据正文中的(3)式，不难有：

$$\delta=(N-1)F_{\mathrm{ML0}}=(N-1)\left(F_{\mathrm{ML}}-\frac{df}{N-1}\right)=T_{ML}-df \quad (\mathrm{C.1})$$

换言之，如果研究者可以将拟合指数“还原”为原始的 χ^2 值，并且知晓模型自由度，那么理论上研究者就可以根据该参数求出非中心参数 δ 的估计值。侯杰泰等人(2004)的书中给出了绝大多数结构方程模型拟合指数的定义，从中不难看出，当知晓信息足够多的情况下，除了少数直接基于模型隐含协方差矩阵定义的拟合指数(如 RMR、SRMR 等)外，其余的拟合指数均可以“还原”为原始的 χ^2 值并进一步计算出非中心参数 δ。但不同拟合指数的使用频率不同，将其换算为非中心参数所需要的额外参数(或者说，换算的“麻烦程度”)也不同，故并不是所有的拟合指数都在实践中被应用，目前较为常用的拟合指数有：RMSEA、MFI、GFI、AGFI 和 CFI。这里特别需要注意的是，目前软件和论文中对 GFI 和 AGFI 的缩写存在一定混用的情况，可能不加区别地将“Goodness of fit index/LISREL GFI” (下称 GFI^{L})和“Unbiased goodness of fit index/Steiger's γ” (下称 GFI^{*}); “Adjusted goodness of fit index” (下称 AGFI^{L})和“Unbiased adjusted goodness of fit index” (下称 AGFI^{*})混用。以 R 为例，R 语言结构方程建模程序包 lavaan 中关于 GFI 的定义采用了 GFI^{L}，而 semPower 包中则采用了 GFI^{*}。特别需要说明的是，MacCallum 法中使用的是 GFI^{*}和 AGFI^{*}。因此若对 lavaan 中得到的 GFI 值进行后验检验力分析，则可能出现不正确的结果。这五种指数与非中心参数换算的公式为：

$$\delta=(N-1)\cdot df\cdot \mathrm{RMSEA}^2 \quad (\mathrm{C.1a})$$

$$\delta=(N-1)\cdot(-2\ln \mathrm{MFI}) \quad (\mathrm{C.1b})$$

$$\delta=(N-1)\frac{p(1-\mathrm{GFI}^*)}{2\mathrm{GFI}^*} \quad (\mathrm{C.1c})$$

$$\delta=(N-1)\frac{p(1-\mathrm{AGFI}^*)df}{p(p+1)-2df(1-\mathrm{AGFI}^*)} \quad (\mathrm{C.1d})$$

$$\delta=(1-\mathrm{CFI})[(N-1)F_{Null}-df_{Null}] \quad (\mathrm{C.1e})$$

chinaXiv:202303.09766v1

ChinaXiv合作期刊

其中, p 为协方差矩阵的秩(即变量维度), F_{Null} 和 df_{Null} 分别为虚模型(Null Model)的拟合函数值和自由度, 其中 $F_{Null}=-\ln|R|$, R 是变量的相关矩阵。

参考文献

侯杰泰, 温忠麟, 成子娟. (2004). *结构方程模型及其应用*. 北京: 教育科学出版社. (pp. 162−165)

◎补充材料 D: 基于 G*Power 的结构方程模型统计检验力分析

本文正文部分在介绍结构方程模型统计检验力分析时主要使用了 R 语言作为软件工具。但 R 语言本身学习成本较高, 对于不熟悉编程的研究者而言不是非常友好。事实上, R 语言并非进行结构方程模型统计检验力分析的唯一途径, 对于那些不熟悉 R 的研究者, 只要研究者本人能较为熟练地使用目前市面上常见的结构方程建模软件(如 LISREL、EQS、AMOS、Mplus 等), 再结合目前学界常用的 G*Power, 即可完成基于 χ^2 分布的结构方程模型统计检验力的分析, 这在一定程度上也能够帮助研究者应对一些不是那么复杂的应用场景。

1. 使用 G*Power 对结构方程模型统计检验力进行分析的原理

正如正文中提到的那样, 基于 χ^2 分布的统计检验力分析的核心包括两部分: ①根据一些方式估计出非中心参数; ②根据非中心参数结合公式(1a)或(1b)算出对应情况下的统计检验力。其中①的估计又可分为基于协方差矩阵理论真值的 Satorra−Saris 法和基于拟合指数的 MacCallum 法, 这部分是需要研究者自己套用公式计算或借助结构方程建模软件进行计算的部分。G*Power 主要针对的是步骤②中的内容, 即在已知零假设成立时对应的非中心参数 δ_0、备择假设成立时对应的非中心参数 δ_1、自由度 df、I 类错误水平 α 时, 根据公式(2a)或(2b)先计算出拒绝域的临界值, 再根据公式(1a)或(1b)计算出统计检验力的过程。

2. G*Power 中的 Generic χ^2 test

不难看出步骤②的核心是计算 χ^2 分布的分位数和单尾累积概率, 在 G*Power 中, 相关功能可以在 Generic χ^2 test 中找到。打开 G*Power 后, 先选择检验族(Test Family)为 χ^2 tests, 而后便可在检验方式(Statistical test)一栏中找到 Generic χ^2 test。G*Power 的 Generic χ^2 test 部分主要提供了四个工具:Compromise, Criterion, Post hoc 和 Sensitivity。其中后三个工具与结构方程模型统计检验力分析有关, 它们的作用分别是: Criterion: 给出统计检验力 $1-\beta$、非中心参数 δ(在 G*Power 中用 λ 表示)和自由度 df, 求出对应的 I 类错误水平 α; Post hoc: 给出非中心参数 δ、I 类错误水平 α 和自由度 df, 求出对应的统计检验力 $1-\beta$; Sensitivity: 给出 I 类错误水平 α、统计检验力 $1-\beta$ 和自由度 df, 求出对应的非中心参数。但特别需要注意的是, G*Power 提供的这些工具都默认零假设成立时对应的非中心参数 $\delta_0 = 0$, 即零假设成立时统计量的分布为中心化 χ^2 分布。在结构方程模型假设检验的经典方法中, 除去等效性检验外的其他方法均能满足 G*Power 的默认条件, 因此可以直接借用 G*Power 进行分析。等效性检验的情况较为复杂, 若等效性检验的 $\delta_1=0$ 时, 可以借用正文 5.6 部分的结论, 构建与其对偶情形, 并且将 α 与 β 颠倒代入求解; 若等效性检验的 $\delta_1>0$ 时, 则需借用中心化 χ^2 分布为“桥”, 建立起两个非中心 χ^2 分布的联系。

3. 具体操作方法与示例

如前所述, 基于 χ^2 分布的统计检验力分析的第一部分主要需要研究者自己套用公式计算或借助结构方程建模软件进行计算, 本文对此步骤不进行详述, 仅简单介绍本步中需要注意的几个要点。①本步中的难点在于选用 Satorra−Saris 法时协方差矩阵理论真值的生成, 特别是基于模型的理论协方差矩阵真值的生成, 这需要研究者对结构方程模型的基本原理有较好的理解水平。目前市面上大部分结构方程建模的教材都对这部分进行了相应的介绍, 系统学习过结构方程建模的研究者通常可以在教材指导下自己完成理论协方差矩阵的生成工作。另外, 生成理论协方差矩阵时需要计算矩阵乘法, 这部分可通过 Excel 的 MMULT 函数完成。②不同软件计算 χ^2 统计量的方法略有不同, 对 Mplus 而言, $\chi^2=NF_{ML}$, 而对

chinaXiv:202303.09766v1

ChinaXiv合作期刊

LISREL、EQS 和 AMOS 而言, $\chi^2 = (N-1)F_{ML}$, 这导致了在使用不同软件进行分析时, 研究者可能需要针对软件的计算方法对公式进行微调。本部分主要展示的是第二部分的内容, 即已知 δ_0、df 和 α, δ_1 和 1−β 已知其一时如何使用 G*Power 求出另一个的对应值。参考前述对 G*Power 中的 Generic χ^2 test 的介绍, 这里主要分三种情况给出具体的操作方法与示例:

(1) $\delta_0 = 0, \delta_1 > 0$

①已知 1−β, 未知 δ_1(对应先验检验力分析): 选择检验族(Test Family)为 χ^2 tests, 检验方式(Statistical test)为 Generic χ^2 test, 检验力分析类型(Type of power analysis)为 Sensitivity, 设定 df、α 和 1−β, 求解出 δ_1, 而后根据 Satorra−Saris 法得到的 F_{ML} 值或 MacCallum 法对应的拟合指数结合补充材料 C 部分的公式推出结果。

例 1: 已知 $df = 51$, 控制 α 水平为 0.05, 求使得当 RMSEA 的真实值不小于 0.08 时正确拒绝零假设的概率不低于 95%的样本量 N_{Plan}。

解: 打开 G*Power, 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Sensitivity, 设定 $df = 51$、α = 0.05、1−β = 0.95, 求得非中心参数为 43.400, 根据公式(C.2a), 求出样本量为 133.965, 向上取整得 $N_{Plan} = 134$ 人。

②已知 δ_1, 未知 1−β(对应后验检验力分析、使用线性插值估计样本量时的先验检验力分析): 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Post hoc, 设定 df、α 和 δ_1, 求解出 1−β。

例 2: 同例 1, 使用设定多个 N 的方式结合线性插值法估计 N_{Plan}。

解: 打开 G*Power, 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Post hoc, 设定 $df = 51$、α = 0.05、根据公式(C.2a)计算出 N 为 51, 101, 151 时对应的非中心参数分别为 16.32, 32.64, 48.96, 分别代入, 求得对应的 1−β 为 0.566, 0.837, 0.975。使用线性插值法求出 N_{Plan} 约为 142 人。

(2) $\delta_0 = k(N-1) > 0, \delta_1 = 0$

①已知 1−β, 未知 δ_0(对应先验检验力分析): 根据正文 5.6 部分的结论, 构建与其对偶情形——$\delta_0' = 0, \delta_1' = k(N-1) > 0$, 而后交换 α 与 β, 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Sensitivity, 设定 df、α 和 1−β, 求解出 δ_1, 进而补充材料 C 部分的公式求得 N。

例 3: 已知 $df = 51$, 控制 α 水平为 0.05, 进行等效性检验, 设定 $RMSEA_0 = 0.08$, 求使得 RMSEA 的真实值为 0 时正确拒绝零假设的概率不低于 80%的样本量 N_{Plan}。

解: 构建对偶情形: $\delta_0' = 0$, 根据公式(C.2a)有 $\delta_1' = 0.3264 \cdot (N-1)$, 此时有 α’ = 0.2, 1−β’ = 0.95。打开 G*Power, 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Sensitivity, 设定 $df = 51$、α = 0.2、1−β = 0.95, 求得非中心参数为 31.564, 进一步求得 N_{Plan} 约为 98 人。

②已知 δ_0, 未知 1−β(对应后验检验力分析、使用线性插值估计样本量时的先验检验力分析): 根据正文 5.6 部分的结论, 构建与其对偶情形——$\delta_0' = 0, \delta_1' = k(N-1) > 0$, 而后交换 α 与 β, 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Criterion, 设定 df、α 和 δ_1, 求解出 1−β。

例 4: 同例 2, 使用设定多个 N 的方式结合线性插值法估计 N_{Plan}。

解: 构建对偶情形: $\delta_0' = 0$, 根据公式(C.2a)有 $\delta_1' = 0.3264 \cdot (N-1)$, 此时有 1−β’ = 0.95。设定 N 分别为 51, 76, 101, 求得对应的 δ_1'为 16.32, 24.48 和 32.64。打开 G*Power, 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Criterion, 设定 $df = 51$、1−β’ = 0.95, 将 δ_1'的三个值分别代入, 求得对应的 α’为 0.614, 0.370, 0.180。使用线性插值法求出 N_{Plan} 约为 99 人。

(3) $\delta_0 = k(N-1) > 0, \delta_1 = q(N-1) > 0, \delta_0 > \delta_1$

此种情形下问题过于复杂, 在此仅给出一种情况下的操作:

已知 δ_1, 未知 1−β(对应后验检验力分析、使用线性插值估计样本量时的先验检验力分析): 注意到, 给定自由度时中心化 χ^2 分布的右尾概率等价于其临界值, 因此可以以此为“桥”连接起两个非中心 χ^2 分布。首先选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Criterion, 设定其统计检验力为等效性检验的 1−α, df, δ_0, 求出中心化 χ^2 分布的右尾概率, 记为 γ。然后选择检验族为 χ^2 tests,

chinaXiv:202303.09766v1

ChinaXiv合作期刊

检验方式为 Generic χ^2 test, 检验力分析类型为 Post hoc, 代入 $\alpha = \gamma$, 非中心参数为 δ_1 和自由度, 求得对应的统计检验力(此时的统计检验力为 β 而非 $1-\beta$)。特别需要注意的是, 这种以中心化 χ^2 分布为“桥”的方法受限于 G*Power 对 α 的计算限制, G*Power 允许的 α 范围是 1e-6 到 1−1e-6, 这也意味着此种情形下 G*Power 的作用是有限的, 对于此种情形, 若非中心 χ^2 分布的自由度足够大时, 也可借用正态分布对非中心 χ^2 分布进行近似从而求出统计检验力的近似值。

例 5: 已知 $df = 51$, 样本量为 301, 设定 $RMSEA_0 = 0.08$ 进行等效性检验, 结果显示 $RMSEA_1 = 0.043$, $p < 0.05$, 求统计检验力。

解: 根据公式(C.2a)计算出 $\delta_0 = 97.92$, $\delta_1 = 28.290$。打开 G*Power, 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Criterion, 设定 $df = 51$、$1-\beta = 1-0.05 = 0.95$, $\lambda = 97.92$, 求得 α 为 1.014e-6。然后选择检验力分析类型为 Post hoc, 设定 $df = 51$、$\lambda = 28.290$、$\alpha = 1.014e\text{-}6$, 求得统计检验力为 0.985。

例 6: 已知 $df = 51$, 样本量为 439, 设定 $RMSEA_0 = 0.08$ 进行等效性检验, 结果显示 $RMSEA_1 = 0.067$, $p < 0.05$, 求统计检验力。

解: 根据公式(C.2a)计算出 $\delta_0 = 142.9632$, $\delta_1 = 100.2753$。打开 G*Power, 选择检验族为 χ^2 tests, 检验方式为 Generic χ^2 test, 检验力分析类型为 Criterion, 设定 $df = 51$、$1-\beta = 1-0.05 = 0.95$, $\lambda = 142.9632$, 求得 α 为 4.064e-012。注意到解得的 α 远超出了 G*Power 允许的范围, 但同时注意到本例中 df 较大, 可以考虑使用正态分布对非中心 χ^2 分布进行近似, 此时读出 α 为 4.064e-012 时对应的中心化 χ^2 分布临界值为 152.944, 同时计算非中心参数为 δ_1、自由度为 51 的非中心 χ^2 分布的均值和标准差, 根据非中心 χ^2 分布的相关结论有:

$$\mu = \delta + df = 100.2753 + 51 = 151.275; \sigma = \sqrt{2df + 4\delta} = \sqrt{2\times 51 + 4\times 100.2753} = 22.430 \qquad (D.1)$$

以正态分布对非中心 χ^2 分布进行近似, 可求得均值为 151.275, 标准差为 22.430 的正态分布割点为 152.944 时的左尾概率约为 0.530, 与之前 B 部分得到的理论值(0.546)还是较为相近的。

◎补充材料 E: 案例中 Monte Carlo 模拟部分各参数估计情况、相对误差以及 95%置信区间覆盖率

表 E-1　模型拟合检验先验检验力的 Monte Carlo 分析

路径	理论值	模拟样本量 $n = 101$						模拟样本量 $n = 201$					
		估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差	估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差
POS → X1	0.870	0.867	0.083	0.084	0.953	0.004	0.018	0.870	0.059	0.058	0.952	0.000	0.009
POS → X2	0.900	0.896	0.082	0.082	0.946	0.005	0.011	0.901	0.058	0.057	0.941	0.001	0.009
POS → X3	0.750	0.745	0.088	0.089	0.947	0.007	0.008	0.752	0.063	0.062	0.952	0.003	0.013
POS → X4	0.550	0.545	0.096	0.096	0.944	0.009	0.006	0.554	0.068	0.068	0.953	0.007	0.002
DES → X5	0.590	0.589	0.099	0.095	0.956	0.001	0.035	0.590	0.070	0.069	0.958	0.001	0.013
DES → X6	0.720	0.719	0.093	0.096	0.939	0.002	0.030	0.725	0.066	0.068	0.942	0.006	0.029
DES → X7	0.580	0.577	0.091	0.096	0.943	0.005	0.049	0.581	0.065	0.066	0.943	0.002	0.017
DES → X8	0.710	0.702	0.094	0.092	0.949	0.011	0.020	0.711	0.067	0.066	0.953	0.001	0.003
DES → X9	0.590	0.587	0.091	0.092	0.947	0.006	0.019	0.593	0.064	0.066	0.946	0.005	0.025
ANG → X10	0.730	0.731	0.093	0.093	0.946	0.001	0.006	0.733	0.066	0.066	0.947	0.005	0.010
ANG → X11	0.780	0.779	0.093	0.094	0.946	0.001	0.013	0.780	0.066	0.069	0.934	0.000	0.040
ANG → X12	0.570	0.574	0.092	0.095	0.937	0.007	0.032	0.572	0.066	0.066	0.948	0.003	0.012
X1 ↔ X1	0.240	0.237	0.057	0.059	0.932	0.013	0.041	0.240	0.040	0.040	0.955	0.001	0.018
X2 ↔ X2	0.190	0.189	0.055	0.057	0.948	0.006	0.029	0.188	0.039	0.039	0.951	0.011	0.003

chinaXiv:202303.09766v1

ChinaXiv合作期刊

续表

路径	理论值	模拟样本量 n = 101						模拟样本量 n = 201					
		估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差	估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差
X3 ↔ X3	0.440	0.437	0.072	0.073	0.932	0.007	0.015	0.438	0.051	0.052	0.944	0.005	0.013
X4 ↔ X4	0.700	0.699	0.104	0.104	0.924	0.002	0.001	0.696	0.073	0.074	0.940	0.006	0.015
X5 ↔ X5	0.650	0.642	0.102	0.106	0.934	0.012	0.032	0.646	0.073	0.075	0.939	0.006	0.038
X6 ↔ X6	0.470	0.466	0.085	0.088	0.947	0.009	0.031	0.461	0.060	0.062	0.924	0.019	0.023
X7 ↔ X7	0.540	0.532	0.086	0.088	0.929	0.014	0.017	0.536	0.062	0.063	0.941	0.007	0.029
X8 ↔ X8	0.490	0.488	0.087	0.088	0.936	0.003	0.013	0.488	0.062	0.061	0.946	0.004	0.014
X9 ↔ X9	0.520	0.516	0.085	0.086	0.937	0.007	0.014	0.519	0.060	0.064	0.933	0.002	0.057
X10 ↔ X10	0.450	0.441	0.086	0.091	0.921	0.021	0.055	0.445	0.062	0.063	0.940	0.011	0.022
X11 ↔ X11	0.390	0.384	0.085	0.087	0.950	0.015	0.018	0.391	0.061	0.062	0.949	0.003	0.012
X12 ↔ X12	0.550	0.539	0.088	0.091	0.929	0.021	0.031	0.546	0.063	0.062	0.939	0.007	0.014
POS ↔ DES	0.230	0.227	0.111	0.117	0.935	0.014	0.050	0.230	0.079	0.080	0.937	0.002	0.009
POS ↔ ANG	0.400	0.398	0.103	0.110	0.929	0.004	0.067	0.401	0.073	0.074	0.953	0.002	0.004
DES ↔ ANG	0.780	0.779	0.071	0.072	0.938	0.001	0.016	0.781	0.050	0.050	0.943	0.002	0.000

表 E-2　模型参数检验先验检验力的 Monte Carlo 分析

路径	理论值	模拟样本量 n = 301						模拟样本量 n = 401					
		估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差	估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差
RES→SRG	**0.180**	**0.179**	**0.057**	**0.057**	**0.951**	**0.006**	**0.001**	**0.181**	**0.049**	**0.048**	**0.949**	**0.001**	**0.026**
PSS→SRG	0.180	0.183	0.057	0.059	0.942	0.017	0.034	0.180	0.049	0.050	0.948	0.002	0.016
ECI→SRG	0.180	0.180	0.057	0.056	0.948	0.002	0.016	0.179	0.049	0.050	0.942	0.006	0.013
STR→SRG	0.038	0.038	0.053	0.053	0.949	0.010	0.009	0.040	0.046	0.046	0.950	0.036	0.000
Ser→SRG	0.038	0.040	0.053	0.053	0.955	0.057	0.001	0.037	0.046	0.046	0.959	0.039	0.012
Age→SRG	0.038	0.039	0.053	0.054	0.949	0.023	0.019	0.039	0.046	0.046	0.953	0.023	0.008
PSS→RES	**0.220**	**0.220**	**0.056**	**0.056**	**0.955**	**0.002**	**0.011**	**0.221**	**0.049**	**0.047**	**0.959**	**0.006**	**0.027**
ECI→RES	0.220	0.222	0.056	0.056	0.949	0.008	0.005	0.219	0.049	0.046	0.960	0.005	0.058
STR→RES	0.047	0.047	0.054	0.057	0.940	0.002	0.049	0.045	0.047	0.048	0.947	0.031	0.018
Ser→RES	0.047	0.046	0.054	0.054	0.949	0.010	0.006	0.046	0.047	0.048	0.947	0.010	0.017
Age→RES	0.047	0.045	0.054	0.054	0.955	0.046	0.006	0.048	0.047	0.046	0.955	0.031	0.022
ECI→PSS	**0.282**	**0.282**	**0.055**	**0.054**	**0.961**	**0.001**	**0.017**	**0.283**	**0.048**	**0.046**	**0.958**	**0.002**	**0.036**
STR→PSS	0.060	0.060	0.055	0.056	0.949	0.010	0.008	0.062	0.048	0.049	0.951	0.038	0.020
Ser→PSS	0.060	0.060	0.055	0.056	0.949	0.005	0.003	0.060	0.048	0.048	0.943	0.005	0.010
Age→PSS	0.060	0.055	0.055	0.059	0.935	0.077	0.062	0.058	0.048	0.048	0.942	0.025	0.003

注: n 为 501 和 601 时的参数估计情况见下页续表, 加粗的表示该路径系数是关键参数; RES: 情绪调节自我效能感, PSS: 领悟社会支持, ECI: 情绪创造力, SRG: 压力后成长, STR: 疫情相关压力, Ser: 疫情严重程度, Age: 年龄

chinaXiv:202303.09766v1

ChinaXiv合作期刊

表 E-2(续)　模型参数检验先验检验力的 Monte Carlo 分析(续表)

路径	理论值	模拟样本量 n = 501						模拟样本量 n = 601					
		估计均值	均值标准误	标准误均值	覆盖率	均值误差	标准误误差	估计均值	均值标准误	估计均值	均值标准误	均值误差	标准误误差
RES→SRG	0.180	0.179	0.044	0.043	0.952	0.008	0.013	0.178	0.040	0.041	0.940	0.011	0.022
PSS→SRG	0.180	0.180	0.044	0.046	0.943	0.003	0.035	0.182	0.040	0.041	0.944	0.009	0.023
ECI→SRG	0.180	0.183	0.044	0.043	0.952	0.015	0.019	0.180	0.040	0.040	0.941	0.001	0.000
STR→SRG	0.038	0.038	0.041	0.040	0.956	0.017	0.019	0.039	0.038	0.038	0.951	0.009	0.006
Ser→SRG	0.038	0.039	0.041	0.041	0.954	0.028	0.002	0.038	0.038	0.039	0.947	0.005	0.031
Age→SRG	0.038	0.041	0.041	0.041	0.957	0.064	0.013	0.039	0.038	0.037	0.947	0.015	0.009
PSS→RES	0.220	0.221	0.044	0.042	0.962	0.003	0.037	0.218	0.040	0.041	0.951	0.010	0.018
ECI→RES	0.220	0.221	0.044	0.042	0.958	0.006	0.045	0.220	0.040	0.039	0.952	0.002	0.029
STR→RES	0.047	0.050	0.042	0.043	0.947	0.067	0.020	0.047	0.038	0.038	0.950	0.005	0.019
Ser→RES	0.047	0.045	0.042	0.041	0.958	0.029	0.010	0.046	0.038	0.039	0.943	0.017	0.009
Age→RES	0.047	0.047	0.042	0.042	0.947	0.013	0.003	0.048	0.038	0.038	0.955	0.038	0.005
ECI→PSS	0.282	0.282	0.043	0.042	0.948	0.001	0.010	0.280	0.039	0.037	0.963	0.008	0.049
STR→PSS	0.060	0.059	0.043	0.042	0.956	0.018	0.023	0.061	0.039	0.040	0.947	0.012	0.011
Ser→PSS	0.060	0.059	0.043	0.043	0.946	0.010	0.010	0.061	0.039	0.041	0.945	0.012	0.041
Age→PSS	0.060	0.060	0.043	0.043	0.945	0.000	0.000	0.059	0.039	0.040	0.948	0.009	0.029

chinaXiv:202303.09766v1

Advances in Psychological Science